8
Historical-Clinical-Risk Management-20 (HCR-20) Violence Risk Assessment Scheme
Rationale, Application, and Empirical Overview

KEVIN S. DOUGLAS and KIM A. REEVES

The *Historical-Clinical-Risk Management-20* (HCR-20) (Webster, Douglas, Eaves, & Hart, 1997) was one of the first violence risk assessment protocols developed under the *Structured Professional Judgment* (SPJ) model of risk assessment, a model that has now been subjected to more than 100 independent empirical studies (Guy, 2008). As described in the sections that follow, the HCR-20 is intended to provide a structured assessment of the risk factors that are present in a given case, the relevance of those risk factors for a given individual's violence risk, and what risk management strategies might be put into place in order to mitigate risk. The *HCR-20 Violence Risk Management Companion Guide* (Douglas, Webster, Hart, Eaves, & Ogloff, 2001) can be used in tandem with the HCR-20 to facilitate identification and conceptualization of risk management plans that link up with the more management-relevant HCR-20 risk factors (as described below). Given its SPJ approach, the HCR-20 differs both from the traditional, unstructured clinical approach to risk assessment and from the actuarial approach. However, as described later in this chapter, it shares features with both.

The HCR-20 has been translated into 16 languages and is used in a large number of correctional, forensic, and psychiatric agencies, systems, and institutions across North America, Europe, South America, Asia, and Australia. There have been more than 50 evaluations of its validity across approximately a dozen countries (Douglas, Guy, Reeves, & Weir, 2008; Guy, 2008). In this chapter, we describe the rationale for the development of the HCR-20, how to use it, its defining features, how its use differs from both unstructured clinical judgment and actuarial prediction, empirical evidence supporting its reliability and validity, and we present a sample case.

Description of Measure

Criterion Being Assessed

The HCR-20 is intended to facilitate assessments of risk for interpersonal violence, which is defined in the HCR-20 manual (Webster et al., 1997) as "actual, attempted, or threatened harm to a person or persons" (p. 24). Note that actual physical harm is not required—*threatened* or *attempted* harm fits the definition of violence. As written in the HCR-20 manual, "[t]hreats of harm must be clear and unambiguous threats (e.g., "I am going to kill you!"), rather than vague statements of hostility" (p. 24). To take an extreme example, if a person were to shoot a gun into a crowded room and strike no one, that is violence. There is simply no meaningful logical difference between attempted and completed violence in terms of the behavior and intention of the perpetrator, *other than* he or she failed. As such, it does not make sense to exclude attempts from the definition of violence.

Property damage or harm to animals is not considered violence, unless carried out in a manner that is intended to cause fear of harm in others (i.e., smashing a chair, or injuring an animal,

Historical (Past)	Clinical (Present)	Risk Management (Future)
H1. Previous Violence	C1. Lack of Insight	R1. Plans Lack Feasibility
H2. Young Age at First Violent Incident	C2. Negative Attitudes	R2. Exposure to Destabilizers
H3. Relationship Instability	C3. Active Symptoms of Major Mental Illness	R3. Lack of Personal Support
H4. Employment Problems	C4. Impulsivity	R4. Noncompliance with Remediation Attempts
H5. Substance Use Problems	C5. Unresponsive to Treatment	R5. Stress
H6. Major Mental Illness		
H7. Psychopathy		
H8. Early Maladjustment		
H9. Personality Disorder		
H10. Prior Supervision Failure		

Figure 8.1 HCR-20 scales and items. (Webster et al., 1997. *HCR-20: Assessing risk for violence*, Version 2. Burnaby, BC, Canada: Simon Fraser University. Reprinted with permission.)

Chapter 11, this volume); description of the measure; general information about violence risk assessment; research on the instrument; user qualifications; scope and purpose; definition of violence; administration principles; item definitions; item coding instructions; and instructions for reaching a final risk decision and informing risk management plans.

The HCR-20 scales and items are presented in Figure 8.1. The three HCR-20 scales are intended to cover past, present, and future. The Historical (H) Scale contains 10 items that cover important past risk factors, the five Clinical (C) Scale risk factors pertain to current or recent functioning, and the five Risk Management (R) Scale factors require ratings about future risk-relevant circumstances. There are 20 risk factors across these three scales—hence the name of the measure: HCR-20. Details about each of these three scales follow.

Historical Scale. The Historical (H) Scale is so named because it indexes a core set of important violence risk factors that may have occurred or transpired at some point in a person's history. It is important to note that these risk factors may very well, and likely often will, have crucial relevance for understanding a person's *current* and *future* risk for violence as well. For this reason, it is essential that evaluators not only diligently record the proper score for each risk factor, but consider the current and future relevance of each risk factor for a person's violence risk. Similarly, it would be a mistake to consider many of these risk factors to be static or unchanging. Whereas a person may always receive a high score on these items (i.e., a person with years of substance abuse) by virtue of the fact that the risk factor was present in the past, evaluators must consider the current and future relevance of each risk factor, as reflected by its current manifestation and course. Therefore, H scale risk factors need to be considered for risk management plans along with C scale and R scale risk factors.

There is no conceptual theme that ties the H scale risk factors together. They are united by temporal period and reflect a summary of what the current literature on violence indicates as core risk factors. However, it may be beneficial to consider these 10 risk factors as falling into several general categories: (1) problems in adjustment or living, (2) problems with mental health, and (3) past antisociality. *Problems in adjustment or living* are captured by H3 (Relationship Instability), H4 (Employment Problems), and H8 (Early Maladjustment). *Problems with mental*

health include H5 (Substance Use Problems), H6 (Major Mental Illness), H7 (Psychopathy), and H9 (Personality Disorder). Finally, *past antisociality* is captured by H1 (Previous Violence), H2 (Young Age at First Violent Incident), and H10 (Prior Supervision Failure).

Clinical Scale.   The Clinical (C) Scale captures *recent* and *current* functioning in important risk-relevant domains. Although most risk factors on the HCR-20 are potentially dynamic (including those on the Historical Scale), those on the C scale focus on phenomena that can change acutely. In some cases, change may be manifest over very short periods of time (e.g., hours or days), whereas in other cases, change may be more gradual (for example, months). There is direct evidence that the C scale items can change over time (Belfrage & Douglas, 2002; Douglas & Belfrage, 2001). This focus on *dynamic risk* is commensurate with an emerging emphasis in the risk assessment field to ensure that risk factors that are highly relevant to intervention and management are featured centrally within assessments (Andrews, Bonta, & Wormith, Chapter 10, this volume; Douglas, 2008; Douglas & Skeem, 2005; Dvoskin & Heilbrun, 2001; Mulvey & Lidz, 1995; Skeem & Mulvey, 2002). Therefore, C scale risk factors ideally should be reevaluated on a regular basis, a point we expand upon below. Corresponding ratings of risk level, and recommended or actual intervention and management strategies, should be adjusted accordingly.

Risk Management Scale.   This section centers on forecasting how individuals will adjust to future circumstances. Although the Historical and Clinical Scales also are relevant to and should figure prominently in the development of risk management plans, the Risk Management (R) Scale is intended to focus evaluators' attention on development of appropriate future risk management plans, to speculate about what impediments to successful management might exist, and how to address such impediments.

The factors captured by the R scale items also can change over time (Belfrage & Douglas, 2002; Douglas & Belfrage, 2001). Therefore, as with the C scale, we recommend that the R scale items be reevaluated on a regular basis, consistent with the person's context. We would recommend that they are reevaluated or monitored whenever the C scale items are reevaluated, because changes in a person's recent functioning can have implications for his or her future functioning and the adequacy of plans.

Contexts and Populations in Which HCR-20 Is and Is Not Appropriate to Use

The HCR-20 is intended to be used with men and women age 18 and above. There is some leeway here; in some cases it may be appropriate to use the HCR-20 with younger persons (say, 16 or 17) if they have been living independently for some time. Similarly, it may make more sense in certain cases *not* to use the HCR-20 with a 19- or 20-year-old (a person living at home and dependent on parents, for instance). It will be up to the evaluator to judge whether the HCR-20 or something like the SAVRY (Borum et al., Chapter 4, this volume) is more appropriate in these cases on the cusp.

As stated in the HCR-20 manual, its use "should be restricted mainly to settings in which there is a high proportion of persons with histories of violence, and a strong suggestion of mental illness or severe personality disorder" (p. 5). The manual continues to explain that this would include forensic and civil mental health settings, and parole and other correctional settings. Research, described in detail later in this chapter, has been conducted and provides support for the use of the HCR-20 in samples of forensic psychiatric patients, civil psychiatric patients, offenders with mental disorders, and general correctional offenders, both for institutional and community violence. There have also been studies on the HCR-20 with young offenders.

More concretely, the HCR-20 can be used in the following settings and contexts: (1) release decision making (from correctional, psychiatric, or forensic facilities); (2) admission decision

making (upon entry to correctional, psychiatric, or forensic facilities); (3) monitoring of risk while a person is incarcerated or institutionalized; and (4) monitoring of risk while a person is under a term of community supervision by correctional, forensic, or psychiatric authorities. The particular manner in which the HCR-20 can be used for such purposes will be described under the "Assessment Procedure: How to Use and Make Decisions with the HCR-20" section.

User Qualifications

The two minimal user qualifications for use of the HCR-20 are "expertise in conduct of individual assessments" (Webster et al., 1997, p. 17) and "expertise in the study of violence" (Webster et al., 1997, p. 17). The former includes training and experience in "interviewing; the administration and interpretation of standardized tests; and the diagnosis of mental disorder" (p. 18). The latter requires evaluators to "be familiar with the professional and research literatures on the nature, causes, and management of violence" (p. 18).

When used for clinical rather than research purposes, evaluators should have "a high level of expertise (e.g., graduate-level university courses or other specialized educational training, supervised field experience) and should have the requisite professional credentials (e.g., registered, licensed, or legally entitled to conduct individual assessments)" (Webster et al., p. 18). When used for research, raters may be well-trained graduate, medical, or even undergraduate students, so long as they are supervised by a person who is fully qualified to administer the HCR-20 under "clinical conditions."

Some items on the HCR-20 require an assessment of mental disorder. As stated in the HCR-20 manual:

[U]sers who lack the requisite credentials to conduct psychodiagnostic assessments are not legally entitled to code these items, except in the following circumstances:

- The items are coded in consultation with or under the supervision of accredited professionals.
- The items are coded by referring to the results of existing psychodiagnostic assessments.
- The items are coded provisionally, with a notation that the coding should be confirmed by an accredited professional.
- The items are omitted altogether, with a notation concerning how their omission may have limited the final judgment of risk. (Webster et al., p. 18. Reprinted with permission.)

Note that the "requisite credentials" may vary across jurisdictions. It is possible for teams of professionals to complete the HCR-20. For instance, a psychiatrist or psychiatric nurse might complete ratings of items pertaining to major mental illness, a psychologist might contribute by assessing items pertaining to personality disorder and psychopathy, and a social worker or probation officer might complete items pertaining to social history and future plans. In this arrangement, it is important that a person with full qualifications oversee the compilation of ratings, and be responsible for finalizing and approving all final ratings and decisions stemming therefrom.

Method of and Rationale for Development

To understand why the HCR-20 (and other SPJ measures) was developed in the way it was, it is necessary to describe the choices available in the risk assessment field prior to the development of the SPJ model—unstructured clinical judgment and actuarial prediction. Of course, the "clinical versus actuarial" prediction issue has been discussed in the social, behavioral, and medical sciences for over 50 years (Meehl, 1954).

Unstructured clinical judgment is based primarily on professional opinion, intuition, and clinical experience. There is absolute discretion in selecting and conceptualizing risk factors, as well as how to integrate them to make decisions (Meehl, 1954). It is an informal and subjective method. Of course, clinical judgment is a crucial part of assessment and intervention. It is geared toward flexibility and relevance to the individual client or examinee. However, in terms of estimating future probabilities of events, or making predictions, there are numerous problems with this approach, as described below. As written by Grove and Meehl (1996), clinical prediction is an "informal, 'in the head,' impressionistic, subjective conclusion, reached (somehow) by a human clinical judge" (p. 294). In terms of *prediction*, "clinical" means that there are no replicable, specified rules for combining or integrating predictive factors. By contrast, the actuarial approach to prediction is "a formal method" that "uses an equation, a formula, a graph, or an actuarial table to arrive at a probability, or expected value, of some outcome" (Grove & Meehl, 1996, p. 294). That is, the defining feature of the term "actuarial" is the specification of replicable, routinized rules for combining or integrating predictive factors.

Both approaches have strengths and both have weaknesses, and the SPJ approach was developed to avoid the weaknesses of both, while trying to achieve the strengths of both. However, it is *not* a combined "clinical–actuarial" approach. Nor is it an "adjusted-actuarial" approach (that is, adjusting the numeric estimate of an actuarial prediction based on intuition or other non-actuarial means). There is no empirical support for the latter. The former is impossible, by definition. That is, the same single decision-making task cannot be both rule-bound and without rules; single decision tasks either are actuarial, or they are not, in which case they are clinical, as defined originally by Meehl (1954). It *would* technically be possible to construct a multistep decision-making *process* that includes both clinical and actuarial components. In fact, this is essentially the process recommended by Monahan et al. (2005; see also Monahan, Chapter 9, this volume) for using their Classification of Violence Risk (COVR), as well as by others (e.g., Doren, 2002). That is, an actuarial estimate is produced by the COVR and then used by a clinician as one piece of information in a discretionary judgment about whether to release or detain a psychiatric patient. Ultimately, such a multistep process is clinical, given that the clinician is left with the task of deciding how to integrate the actuarial estimate with other information.

Strengths and Weaknesses of Unstructured Clinical Discretion.   Strengths of the clinical approach include responsivity and sensitivity to the individual case. This can enhance case conceptualization and working with people individually to derive risk management plans to reduce violence. Further, clinical decision making is highly flexible and widely applicable.

However, a purely unstructured clinical approach suffers serious weaknesses, given the absence of structure. This includes potentially low reliability, because different clinicians may combine or weight information differently for the same people. Further, decision-makers may not consider factors that actually relate to violent behavior, or consider other factors that do not relate to violence. Not only may there be inconsistency across raters, there also might be inconsistency within raters but across cases. As such, it is likely that such decisions will be of lower reliability and predictive validity than otherwise possible.

Because there are no rules for what risk factors to consider or how to use them in a risk assessment, and an evaluator does not specify such risk factors or how they were used, there is little transparency in unstructured clinical decision making, which is a problematic feature in legal settings, because it makes the later review of such decisions difficult. Hence, such decisions potentially jeopardize the rights of those about whom they are made, as well as thwart attempts at continuity of care. Other potential problems include susceptibility to decisional biases and heuristics. As such, we attempted to avoid these problems in the development of the HCR-20, as described in the sections that follow, yet retained some of the positive features of a clinical approach.

Strengths and Weaknesses of Actuarial Decision Making.   A strength of actuarial prediction is that, being bound by combinatory rules, it facilitates inter-rater reliability and predictive validity, especially in comparison with unstructured approaches. In addition, because the rules used to combine items in order to reach risk decisions are specified *a priori*, actuarial predictions tend to be transparent, which is a benefit in legal contexts. In developing the HCR-20, we were interested in ensuring the features of reliability, validity, and transparency—just not through actuarial means, for reasons described next.

The predictive properties of most actuarial models used in the violence risk assessment field tend to be optimized within a development sample—often just a single sample is used, and the results are not cross-validated prior to use (Minnesota Sex Offender Screening Tool, MnSOST; Violence Risk Appraisal Guide, VRAG). That is, researchers measure some group of patients or offenders on some set of risk factors, observe who is and is not violent during some specified time window, and determine what risk factors are most strongly associated with violence in the extant sample. Typically, actuarial risk assessment instruments, such as those described in this book (VRAG; Sex Offender Risk Appraisal Guide, SORAG; COVR; Static-99; Level of Service Inventory-Revised, LSI-R), produce *estimated* probabilities of violence over some future time for persons who fall into certain score ranges on the test. Some, but not all, weight the individual risk factors on the risk assessment instrument according to their predictive validity in development (and, ideally, cross-validation) samples.

Although the promise of high reliability and accurate numeric estimates of individuals' probabilities of future violence is alluring, there are a number of vulnerabilities of the actuarial approach that, as of yet, have not consistently been demonstrated to have been overcome in the risk assessment field. Most actuarial procedures select risk factors because of demonstrated statistical associations with outcome (violence) within a given sample or samples. While such variables are related to violence in those samples, there is no guarantee that (1) they would have been so in other samples and (2) other risk factors would not have been associated with violence in other samples. Strict actuarial approaches (e.g., VRAG, see this volume) do not permit consideration of any variables that are not included on the instrument, which presumes that the original research that developed the instrument considered all potentially relevant risk factors, and all potentially relevant risk factors are in fact contained within the instrument. If actuarial procedures also weight variables, the presumption is that the same variables would have been weighted comparably in other samples, and that such weights apply equally to all persons in the development sample and any other sample in which the instrument is intended for use. Further, actuarial approaches presume that the observed probabilities of violence in development samples will apply to different samples.

The potential problems with this approach lie in the fact that actuarial approaches are *sample dependent*. That is, (1) the variables that are identified by whatever statistical procedures that are used to select them, (2) the weights given to such variables, and (3) the resulting observed probabilities of violence are all dependent on a myriad of sample-specific characteristics. Such characteristics include, but are not limited to, the variables chosen for study in the first place, the way in which such variables are defined and measured, the reliability of measurement, the demographics of participants, follow-up length, method used to detect violence, definition of violence, and the type of statistical procedures used. Given this sample dependence, any numeric estimates of risk for violence are subject to change if tested in new samples. This potential change has several possible implications: (1) estimates provided by actuarial instruments may, or may not, apply outside of development samples, and this cannot be known unless tested empirically; (2) the strength of prediction may diminish in new samples, to the extent that such new samples differ from the development sample (in fact, even if they are highly similar, predictive strength may diminish because the original estimates also capitalized on chance associations in the

data). Hence, unless numeric estimates are demonstrated to be stable over numerous samples of intended application, it is risky to assume that they will in fact be stable. The basic problem here is a potential lack of robustness and generalizability of risk estimates.

We illustrate some of these problems. As a simple demonstration of the "if you don't measure it, it can't be on your instrument" problem, note that the Static-99, (see Chapter 12), one of the best validated sexual violence risk assessment measures, does not include "sexual deviation" as a risk factor, despite its strength in the empirical literature as a robust risk factor (Laws & O'Donohue, 2008). Other actuarial instruments fail to include risk factors that have broad support in the literature. For example, although the VRAG contains many common risk factors with a good deal of empirical support (i.e., psychopathy, alcohol use problems, separation from parents), there are a number of other risk factors (e.g., anger, previous violence, treatment non-compliance, stress) with empirical support across numerous samples that are not included for the simple reason that they either were not tested as candidate risk factors, or they were tested, but they failed to demonstrate independent statistical associations with violence *in the development sample.* This does not mean that such factors as sexual deviance or anger are irrelevant to violence, just that they were not included on some actuarial instruments because the authors either chose not to test them, they were not related to violence in the development research, or the data were not collected or available for review.

If actuarial instruments weight variables, the presumption is that the weights (1) strengthen the prediction and (2) apply equally to all people (that, say, alcohol problems are always 2.5 times more important than, say, child abuse, in determining future risk of violence). Grann and Långström (2007) tested the first assumption in a sample of 404 Swedish forensic psychiatric patients whom they followed in the community for approximately two years after forensic evaluation. They tested the accuracy of four types of weighting procedures (of the HCR-20 H scale) against unit weighting (all variables contributing equal weight to the predictive equation—the scoring option used for the HCR-20). Importantly, they first derived predictive algorithms on a subsample of cases and then tested the predictive accuracy of those algorithms in a new (cross-validation) subsample of cases. Such a procedure is vital to minimize the artificial inflation of predictive accuracy that many actuarial instruments enjoy because they are based on developmental samples, and not cross-validation samples (Harris, Rice, & Quinsey, 1993; Quinsey, Harris, Rice, & Cormier, 1998, 2006). Grann and Långström reported that the more complex the weighting procedure, the greater the diminishment in predictive accuracy upon cross-validation, with unit weighting achieving the strongest predictive accuracy. This finding confirms what Dawes (1979) asserted years ago—that unit weighting is just as accurate as cross-validated weighting procedures. He eloquently referred to this observation as the "robust beauty of improper [unweighted] linear models" (p. 571).

Blair, Marcus, and Boccaccini (2008) demonstrated the shrinkage that may occur upon cross-validation in a meta-analysis of the VRAG, SORAG, and Static-99. They coded correlational effect sizes for these instruments and then divided them into development samples and cross-validation samples. They further divided cross-validation samples into those conducted by the authors of the instruments and those conducted by independent authors unaffiliated with the instruments. They reported shrinkage for each instrument. For example, the effect size for the VRAG in its development sample was .44; it was .36 in cross-validation studies that included at least one of its authors; it was .30 for cross-validation studies that did not include any of its authors. For the SORAG, these estimates were .46, .35, and .29, respectively. For the Static-99, the development sample correlation was .33, and the independent cross-validation correlation was .27 (there were no cross-validations by the authors of the Static-99). Across all measures, development correlations shrank from .39 to .36 to .28.

The implication of such findings is that the numeric estimates of violence may not hold up upon cross-validation. The more precise the estimate (that is, point estimates such as "62%" as opposed to ranges of estimates), the greater the risk that it will not actually apply in new samples. Mills, Jones, and Kroner (2005), in a sample of 209 offenders, tested the robustness of LSI-R and VRAG estimates of recidivism by comparing the observed recidivism rates of reoffending (LSI-R) or violence (VRAG) to the estimates offered by each instrument, which were based on development samples. They reported that their data:

> . . . does not support the use of the initial validation probability bins of either instrument with our sample. The VRAG nine-bin system has the greater problem given the presence of probability reversals through the bins in conjunction with dissimilar probabilities associated with the bins. (p. 579)

In the study conducted by Mills et al. (2005), some of the VRAG bins associated with higher estimates of future violence compared to other (lower estimate) bins actually produced lower observed rates of violence. For instance, the ninth of nine VRAG bins produces an estimate of 100% probability of future violence, according to the authors (Quinsey et al., 2006). In Mills et al., the observed recidivism rate was just 33%. This observed rate of recidivism was in fact lower than the observed rates of recidivism in bins seven and eight, which, according to the VRAG, should be associated with lower, not higher, rates of future violence. What is the "correct" estimate of violence, then, for people who score in the ninth (or fifth, or second, etc.) bin of the VRAG? The answer is that there is no correct numeric estimate of future violence—such observed estimates are prone to bounce around across samples. Hence, it is possible or even likely that offering an ostensibly precise numeric estimate of recidivism will simply be misleading.

While it generally would be safe to assume that a person who scores higher on some actuarial instrument than another person would be at greater risk for violence, as demonstrated by meta-analytic linear associations of moderate magnitude between such instruments and violence (Blair et al., 2008), it is *not* safe to declare what the estimated numeric probability of violence of either person would be. In fact, it does not appear that risk assessment technology has yet advanced to a point where this is possible.

There are a couple of other potential limitations to some actuarial approaches. Many such approaches tend to be less relevant to risk management and treatment because they include mainly historical, static variables (an exception is the LS system, see Chapter 5 and Chapter 10, this volume), rather than dynamic risk factors that inform decisions about risk state as opposed to risk status (Douglas & Skeem, 2005). A corollary effect of this feature is that risk estimates can be unalterable, or "frozen in time." Regardless of changes in the risk management plan, or intervention successes, a person's risk will remain the same. Given that certain intervention approaches are effective in the reduction of crime and violence (Andrews & Bonta, 2004; Douglas, Nicholls, & Brink, 2009; Dowden & Andrews, 2000), such a static presumption is unfounded.

Rationale for Development of the HCR-20 (as an Instantiation of the SPJ Model).   To combat the weakness of both the clinical and actuarial approaches, as outlined above, yet to try to incorporate their respective strengths, we aimed for the following characteristics in developing the HCR-20.

We used a *logical* or *rational* item selection method to foster generalizability and comprehensiveness of the risk factors on the HCR-20. This approach is in contrast to the empirical item selection approach used by most actuarial methods, the weaknesses of which were outlined above. Logical item selection involves a thorough review of scientific, theoretical, and professional literatures on some topic (here, violence) and the selection of risk factors with support across numerous samples and contexts. The goal is to avoid the omission of important risk

factors or inclusion of unimportant risk factors because of chance associations in a development sample. In principle, because this approach is not sample dependent, risk factors should be generalizable across a variety of settings.

Also in contrast with actuarial approaches, the HCR-20 and other SPJ instruments do not use (1) score cutoffs to determine risk level, (2) numeric estimates of future risk for violence. In our view, such features of risk assessment instruments may perform well in development samples, but have great difficulty in achieving consistency upon cross-validation, as discussed above. Rather, as described in more detail below, the HCR-20 requires clinicians to arrive at a nonnumeric, categorical risk estimate of Low, Moderate, or High risk, based on (1) the presence of risk factors, (2) their relevance of risk in the instant case, and (3) the degree of management, supervision, or intervention required to mitigate risk. This feature differs sharply from actuarial approaches and has been criticized because it allows discretion at the risk decision stage (Quinsey et al., 2006). However, as described in the research section of this chapter, empirical evidence supports this feature of SPJ instruments.

The HCR-20 uses unit-weighting, rather than presuming that group-based statistically weighted and optimized risk factors will apply equally across samples and to all persons. Such an approach, as discussed above, actually may fare worse upon cross-validation compared to simple unit weighting (Grann & Långström, 2007). We do not assume that a given risk factor (say, substance use problems) will be equally important for all persons who have it. For some, it may be highly relevant to risk for violence. For others, it may not. We require users of the HCR-20 to determine for whom it is relevant, and for whom it is not. The rationale for this approach was to promote consideration by users of both the nomothetic level of analysis (risk factors supported through empirical studies of research samples) and the idiographic level (the relevance and manifestation of such nomothetically supported risk factors at the individual level).

The HCR-20, like other SPJ instruments, includes dynamic risk factors in order to facilitate the development of risk management plans. As described elsewhere (Douglas & Skeem, 2005; Skeem & Mulvey, 2002), estimating the risk level of a person may help in terms of determining the appropriate *intensity* of management, but it does little to determine the *type* of management that will reduce risk. For this reason, risk factors on the HCR-20 (particularly the C and R items) should be reevaluated, and management plans should be developed to mitigate them. Relevance to risk management and intervention is one of the strengths of the clinical approach that the HCR-20 attempted to incorporate.

However, as noted above, a purely unstructured approach is likely to produce lower reliability and validity than actuarial approaches, given that the latter have formal rules for the combination of risk factors. Therefore, we included a certain amount of structure into the HCR-20 in order to foster reliability and validity, features often associated with though not defining of the actuarial approach. Structure is found on the HCR-20 by the (1) inclusion of a fixed set of risk factors, (2) operational definitions of risk factors, (3) scoring or coding procedures for risk factors, and (4) direction for how to reach a final decision about risk based on consideration of the risk factors present and relevant in an individual case. These structural elements are described in more detail in the next section.

Assessment Procedure: How to Use and Make Decisions With the HCR-20

Administration Procedures

We recommend the following procedures and steps to complete an HCR-20 evaluation. Evaluators should gather information from as many sources as necessary in the context in which they work.

General categories of sources of information include: (1) relevant files and records; (2) interview with the examinee; (3) psychological testing and assessment procedures; (4) interviews with persons who have knowledge of the individual being evaluated; and (5) direct observation. The general principle we endorse is that evaluators should obtain as much relevant information as necessary to minimize the risk that important details about an examinee are missed.

Sources of Information.   It is vital to secure as much file-based information as necessary in order to complete the HCR-20. This includes but is not limited to nursing notes, social histories, prior mental health evaluations, police reports, arrest records, vocational and educational evaluations, psychological test results, notes from past programming, treatment, or supervision experiences, and records detailing the examinee's adjustment while living in the community under supervision or in an institution.

We recommend, whenever possible, an interview of the person being evaluated. This is consistent with most ethical and practice codes of conduct pertaining to mental health and other human service professions. We do not have a standardized interview protocol, because our goal for the HCR-20 is that it can be integrated into existing practice regimes. The areas that should be covered in such an interview are similar to the areas that should be covered in a file review. An interview to complete the HCR-20 should not require evaluators to ask many more questions than they otherwise would in a comprehensive, thorough psychosocial interview intended to provide a full picture of an individual's history, risk factors, mental health status, past crime and violence, previous treatment and supervision experiences, current functioning, personality, and interpersonal relations. The interview is especially important for assessing current functioning captured by the Clinical Scale, as well as personality and attitudinal characteristics.

Formal psychological testing and assessment procedures will not be required in every case, or perhaps even in the majority of cases. One exception is the use of the Psychopathy Checklist–Revised (PCL-R) or Psychopathy Checklist-Revised: Screening Version (PCL:SV) for use in coding item H7 (Psychopathy). As discussed below, if evaluators must omit an item, including H7, they should state the limits that this might place on the accuracy and comprehensiveness of their evaluation. We leave it to the discretion of evaluators to decide whether such testing is required in a given case, and for what sorts of issues. Such testing could include, inter alia, intelligence and cognitive testing, including neuropsychological testing; semistructured assessment protocols for mental illness and personality disorder; personality assessment using broad-band personality measures; focused tests or assessments of individual constructs (i.e., insight; treatment motivation). We would recommend such testing for the purpose of completing the HCR-20 if, in the judgment of the evaluator, there was one or more areas on the HCR-20 that could not be completed without it.

Where possible and permissible, we recommend that evaluators conduct interviews, even by phone, with others who know the person who is being evaluated. Such third parties could be other professionals, friends, family, or coworkers. Of course, evaluators are expected to obtain any ethical or legal waivers to speak to third parties.

Although interviews clearly are an opportunity for observation, we recommend that, if possible, evaluators observe individuals in less formal contexts as well (that is, behavior on the ward; interaction with others). We realize that direct observation is only possible in certain contexts. We would encourage evaluators to take advantage of such possibility when it exists.

Evaluation Steps

We recommend the following steps be followed in an HCR-20 evaluation:

1. Gather as much information about the individual as is necessary using the five sources of information above. At a minimum, this should include an interview and review of case records.
2. Collate this information as it pertains to each of the HCR-20 risk factors. We recommend, when scoring HCR-20 risk factors, that evaluators consider the evidence both for and against the risk factors, so that the scoring and formulation process is as balanced as possible.
3. In addition to assigning scores of 0 (absent), 1 (possibly or partially present), or 2 (definitely present) to each risk factor, evaluators must engage in two further tasks in order to facilitate case formulation and risk management planning:
   a. Description of the *manifestation* of each risk factor that is present for the case at hand. It is important for evaluators to go beyond the mere noting of a risk factor, and to describe how that risk factor manifests in the individual being evaluated.
   b. Description of the *relevance* of each risk factor that is present to the risk of violence posed by the individual being evaluated. Even if many risk factors are present, evaluators should determine which are most concerning for an individual's risk for violence and, hence, most necessary to manage. In part, this can be done by consideration of the role that each risk factor has played in a person's previous violence.
4. Delineation of the treatment, supervision, or management strategies that logically flow from the risk factors that were identified, taking into account those that are considered most relevant to a person's risk for violence.
5. Determination of whether there are any items or other elements of the decision-making process that had to be omitted because of incomplete or conflicting information. If so, we strongly recommend revisiting those areas of the evaluation in order to maximize the chance that the evaluation is based on complete and valid information about an individual. If any risk factors cannot be coded and hence contribute to case formulation, evaluators should note whatever limits this missing information has on their judgments.
6. Determine whether reevaluation will be needed in the future and, if so, how soon.
7. Statement of whether the individual is considered to be Low, Moderate, or High risk. The complete meaning of these nonnumeric risk categories is described below. It is important to note that their meaning is *not* limited solely to the likelihood of future violence, although it has been inaccurately described as such by some critics (Quinsey et al., 2006).
   a. *High risk* should be applied to cases (1) with many relevant violent risk factors present, or (2) that require frequent, intensive, or highly restrictive supervision, monitoring, management, or intervention in order to stem violence risk. In general, evaluators should make a decision of High risk if they believe, based on the number and relevance of risk factors that are present, and the associated degree of intervention, supervision, monitoring, or management required to mitigate risk, that a person will likely be violent in the future if no appropriate risk management plans are enacted.
   b. *Low risk* should be reserved for cases (1) in which there are few relevant violent risk factors present, or (2) that require minimal or no supervision, monitoring, management, or intervention in order to stem violence risk.
   c. *Moderate risk* should be applied to cases which are neither high nor low risk, as defined above.

It is important to point out that the making of a rating of Low, Moderate, or High risk is almost peripheral to the main task at hand—identifying and managing risk factors. If using

the *HCR-20*, it is *not* acceptable simply to score the items and provide a summary risk rating of Low, Moderate, or High risk. The summary risk rating is merely a shorthand way to note that the person is considered likely to be violent and in need of services. It is a vehicle through which evaluators express their judgments about case prioritization, that is, which persons have a high number of risk factors that require a high degree of intervention. The nature of such intervention must also be specified. The most important part of the risk assessment is to specify what those services are, and to try to put them into effect. As such, decisions of High risk and, to a lesser extent, Moderate or Low risk must be accompanied by (1) a rationale for the decision based on the evidence at hand (number and relevance of risk factors) and (2) the delineation of the risk management, supervision, monitoring, or intervention strategies required to mitigate risk.

We are aware that some evaluators would like both numeric cutoffs and associated numeric estimates of future violence. For reasons discussed above, we do not believe that the current state of the science permits either of these features in the risk assessment field, at least as applied to individuals rather than samples. Further, as reviewed below, the empirical evaluation of the SPJ system of nonnumeric categorical risk categories shows that they are as or more accurate in determining who will be violent in the future compared to actuarial approaches.

Team Evaluations

In some settings, responsibility for conducting assessments may be divided among several different professionals. For example, a psychologist or psychiatrist may assess major mental illness; a psychologist may assess personality or intellectual functioning; and a nurse, social worker, or probation officer may assess release plans. It is acceptable—and may even be desirable—for the HCR-20 to be completed on the basis of reports submitted by several independent professionals. However, a fully qualified user must assume responsibility for integrating the various components, and for making and documenting opinions.

Time Frame for Decisions About Risk for Violence

Historical Scale. The scoring time frame for all Historical items is lifetime, unless otherwise specified within the item. That is, has the risk factor been present *at any time in a person's life?* For example, if a person seriously abused substances 15 years ago but has been abstinent since, the evaluator would indicate that the risk factor is present. *However,* the evaluator might also indicate that the risk factor is not relevant to the person's current or future risk for violence, and it may play a small role in case conceptualization.

Although persons may always score high on H scale items by virtue of their past experiences or behaviors, evaluators should be aware that the coding of H scale items may need to be updated. This may occur if (1) a risk factor that previously was absent (or possibly/partially present) becomes present due to recent behavior (that is, if a person starts to abuse substances and had not done so in the past), or (2) a risk factor has been "dormant" for some time, in which case a score might be reduced from a score of 2 to a score of 1. H scale items can also be updated to correct past coding errors.

Clinical Scale. C scale risk factors ideally should be reevaluated on a regular basis. We recognize that some evaluators will be in a position to rate these risk factors only once. If other professionals then assume responsibility for the management of a person, we would recommend that these factors are reevaluated once the next evaluator has assumed responsibility for the case. In some situations, we realize, the C scale will be used for a single time-point evaluation because circumstances do not allow for reevaluation. If at all possible, however, these risk factors should be tracked over time, and corresponding ratings of risk level, and recommended or actual intervention and management strategies, should be adjusted accordingly.

The C items are intended to index recent and current functioning. However, there simply is no uniform evaluation window that can apply across settings. Indeed, for first-time evaluations, some clients may never have been evaluated previously, or if they have, evaluation results may not be known or available to the current evaluator. We offer some guiding principles to determine the best evaluation window:

1. If a person is being monitored regularly by a professional (i.e., through scheduled appointments), we recommend that the C scale items are updated at each appointment or session (be they weekly, monthly, biyearly). In such situations, the pertinent evaluation window should be the time since the person was last seen.
2. If a person is being evaluated for the first time by a particular professional, the evaluation window should be the previous six months. If the evaluator does not have information about the person that covers the past six months, then the evaluation window should cover as much of the past six months as possible, for as many of the risk factors as possible.

Evaluators also must decide what the time between reevaluations of the C scale items should be. As with the evaluation window, there is no single evaluation interval that makes sense in all contexts. We offer some guiding principles:

1. Higher-risk individuals should be reevaluated more frequently than lower-risk individuals. For higher-risk individuals, evaluators may decide that monthly, weekly, or in some cases, even daily, reevaluations are needed. For lower-risk individuals, evaluators may consider six months to be an appropriate reevaluation interval.
2. We recommend that reevaluation intervals not be longer than six months, if possible, unless a person has shown stable functioning and no violent behavior for at least twelve consecutive months. In such cases, annual reevaluations may be preferable.
3. Notwithstanding these general principles for determining the length of the evaluation interval, a reevaluation should take place under the following circumstances:
   a. There have been notable recent changes in a person's functioning.
   b. The person has engaged in recent violence.
   c. A transition is being considered or will occur (i.e., change in security or supervision level; release from institution).

In addition to rating whether the C items were present *at any time in the evaluation window,* evaluators can indicate whether the C item risk factors *currently* are present. This means that evaluators should not necessarily "take the average" of the risk factor across the evaluation period. For example, if during a six-month evaluation window a person had active symptoms of psychosis in the first month, but has not shown them since, that person should receive a score for the pertinent item, which indicates that the risk factor was present at some point during the evaluation window. The evaluator also should indicate that the risk factor is not *currently* present. By using this approach, the evaluator will know the *trajectory* of each risk factor over the evaluation window (increasing, decreasing, or stable). The evaluator may also choose to determine whether the risk factor has been fluctuating during the evaluation window, which will provide more information about the nature of the risk factor for a given person.

Risk Management Scale.   Our recommendation for evaluation intervals for the R scale are similar to those for the C scale, as follows:

Historical-Clinical-Risk Management-20 (HCR-20) Violence Risk Assessment Scheme • **161**

1. Higher-risk individuals should be reevaluated more frequently than lower-risk individuals. For higher-risk individuals, evaluators may decide that monthly, weekly, or, in some cases, even daily, reevaluations are needed. For lower-risk individuals, evaluators may consider six months to be an appropriate reevaluation interval.
2. We recommend that reevaluation intervals should not be longer than six months, if possible, unless a person has shown stable functioning and no violent behavior for at least twelve consecutive months. In such cases, annual reevaluations may be preferable.
3. Notwithstanding these general principles for determining the length of the evaluation interval, a reevaluation should take place under the following circumstances:
   a. There have been notable recent changes in a person's functioning.
   b. The person has engaged in recent violence.
   c. A transition is being considered or will occur (i.e., change in security or supervision level; release from institution).

Whereas for the C scale, evaluators must make ratings based on some logically defined recent period of time, for the R scale evaluators must make ratings based upon some logically defined *future* period of time. Essentially, evaluators should specify the approximate future time frame for which they are making their ratings. We recommend that this time period mirrors the evaluation interval. That is, if a high-risk person is being seen monthly by a professional, that professional should make R scale ratings for the month to come. If a person were more stable and being reevaluated semiannually or even annually, then the R scale ratings should project forward to cover what is foreseeable over the next six or twelve months.

Finally, R scale ratings can be completed for people who are residing within institutions and will be for some time, people for whom release is near, or people who already are residing in the community. Two types of R scale ratings can be made: *Institutional* ("In") or *Community* ("Out"). Under an "In" evaluation, evaluators rate the items for the period of institutionalization. For an "Out" evaluation conducted on people who currently are residing within an institution, evaluators rate the items based on discharge or release plans and planning, *as if the person were to be released right now.* That is, if the person were released right now, are his or her plans adequate to manage risk in the community? For persons residing in the community, evaluators should rate the items based on the assumption that they will continue to live in the community, unless other arrangements are known to the evaluator.

Overview of Research

Narrative (Douglas, Guy et al., 2008) and meta-analytic (Guy, 2008) reviews indicate that there have been over 50 studies on the HCR-20, and hence it would be impractical to review each of these individually. As such, we discuss the research on the HCR-20 using a "quasi-meta-analytic" approach. That is, for the most part, we report ranges and central tendencies of effect sizes across studies in any given research category (e.g., inter-rater reliability; predictive validity), and supplement this approach with discussion of representative individual studies.

Douglas and Kropp (2002) outlined an approach to the empirical evaluation of any SPJ measure. It is important first to establish inter-rater reliability and predictive validity of the ratings of risk factors. This demonstrates that the risk factors, as defined on a given instrument, can be agreed upon and, in turn, relate to violence (hence supporting their inclusion on SPJ instruments as risk factors). Second, it is important to test whether the judgments of evaluators who use SPJ instruments demonstrate both inter-rater reliability and predictive validity. This second feature of reliability and validity tests the nonactuarial judgments of risk that are required by SPJ instruments and are to be used in clinical practice. It is unique to SPJ instruments relative to actuarial instruments, since most actuarial instruments produce risk decisions purely

on the basis of test results and do not permit evaluators to use their judgment in terms of offering risk estimates (although, as discussed above, proponents of some actuarial approaches advocate that actuarial tools be used in such a way). As such, our discussion of reliability and validity will provide a clear demarcation between numeric HCR-20 *scores*, and HCR-20 facilitated *judgments*.

Reliability

The most important type of reliability for the HCR-20 is inter-rater reliability. Structural reliability is of secondary importance, because the HCR-20 is not a measure of a psychological construct (e.g., depression; anger; psychopathy), and hence its items are not expected to "hang together" as should measures of psychological constructs. Nonetheless, there have been a few reports of internal consistency, which we review before focusing on inter-rater reliability.

Internal Consistency.   Two studies have analyzed the internal consistency of the HCR-20. First, Belfrage (1998) investigated the Swedish translation of Version 1 of the HCR-20 (Webster, Eaves, Douglas, & Wintrup, 1995). Six clinicians scored the HCR-20 on the same 43 forensic patients and compared these scores to the patients' index offense and criminal records. Internal consistency, using Cronbach's alpha, for the HCR-20 total score was high at .95. Acceptably high alphas were obtained for the H, C, and R subscales, respectively (.96, .89, .85). In the second study of internal consistency of the HCR-20, Dunbar, Quinones, and Crevecoeur (2005) investigated Version 2 of the HCR-20 with 204 male offenders convicted of hate crimes in the United States. Similar to Belfrage (1998), all alpha coefficients were high (.94 for H; .90 for C; .95 for R).

Inter-Rater Reliability.   The inter-rater reliability (IRR) of the HCR-20 has been investigated extensively. Most investigators have reported the IRR of the total or scale scores, with fewer having examined the IRR of the summary risk ratings of low, moderate, or high risk. We review these in turn. We also point out that studies have used different indices of IRR. Most have reported intraclass correlation coefficients (ICC), typically single measure. However, some studies have reported other indices (e.g., Pearson correlation coefficient; Kendall's tau). The reported ranges and central tendencies include all reliability coefficients, regardless of type.

Overall, across 36 studies (13 unpublished and 23 published) the IRR of the HCR-20 total numeric score is good to excellent.* For instance, across 25 studies, the median reliability coefficient for the HCR-20 total score is excellent, at .85.† Although IRR has ranged from .67 (Hildebrand, Hesper, Spreen, & Nijman, 2005) to .95 (McDermott, Edens, Quanbeck, Busse, & Scott, 2008), the majority of studies report IRR coefficients of .80 or greater. One study (de Vogel & de Ruiter, 2005) provided two reliability coefficients, one for males (ICC = .77) and one for females (ICC = .75).

---

* Cicchetti and Sparrow (1981), referring to κ and $κ_w$ (which are equivalent to ICC) defined reliability indices below .40 as "poor," .40 to .59 as "fair," .60 to .74 as "good," and .75 or above as "excellent." These authors reported that these categories were consistent with those proposed by several other commentators. Landis and Koch (1977) offered the following scheme: below 0.00 (poor), .00 to .20 (slight), .21 to .40 (fair), .41 to .60 (moderate), .61 to .80 (substantial), and .81 to 1.00 (almost perfect).

† Some studies excluded the R subscale when coding the HCR-20 and hence reported the reliability of an "HC-15" composite. For instance, Grevatt, Thomas-Peter, and Hughes (2004) reported a Kendall's tau of .74 for this composite in a sample of 44 male forensic inpatients. Hill, Habermann, Klusmann, Berner, and Briken (2008) reported an intraclass coefficient (ICC) of .77 among a German sample of 166 male sexual homicide perpetrators.

For the H scale (24 studies), the median IRR value was .86, with a range from .58 (Dunbar et al., 2005) to .97 (Doyle & Dolan, 2006).* As with the HCR-20 total score, the majority of studies reported reliability coefficients of .80 or greater. IRR for the C scale is somewhat lower than for the Total or H scales, with a median across 20 studies of .74. The range extended from .55 (de Vogel & de Ruiter, 2005) to .95 (Howard, 2007), with the majority being .70 or greater. Across 19 studies, the median IRR coefficient for the R subscale was .68, with a range from .47 (Douglas, Ogloff, & Hart, 2003) to .98 (Howard, 2007). There was a split in the IRR coefficients, with a cluster between .51 and .69 and a second cluster between .81 and .98.

The lower reliabilities for the C and R subscales are most likely due to several factors. First, they likely require more judgment to score than the H scale items. Second, a significant proportion of studies do not rely upon mental health professionals as raters, who may be more adept at recognizing such factors such as insight or psychotic symptoms. Third, a number of studies have been "file only" studies, not including an interview. It is possible that without the benefit of an interview, ratings may suffer because there is (1) no chance for observation of current or recent functioning, and (2) no opportunity to tailor the assessment process to the type of information required to rate the HCR-20. Despite this, it is noted that the majority of studies report good or better IRR for these two subscales.

Although the summary risk rating (low, moderate, or high risk) is an important feature of the SPJ model, there are unfortunately few studies that have investigated its reliability. Because the summary risk rating essentially is a single item, its reliability can be expected to be lower than the reliability of a multi-item composite. The median IRR coefficient (ICC) is .65 across nine values (drawn from five studies), with a range of .41 (Douglas, Yeomans, & Boer, 2005) to .76 (de Vogel, de Ruiter, Hildebrand, Bos, & van de Ven, 2004). Because of the small number of investigations, each will be described in more depth. First, in a study of forensic psychiatric patients, de Vogel and de Ruiter (2005) coded the Dutch version of the HCR-20 (from file information only) on 42 female forensic psychiatric patients with personality disorders and a matched sample of 42 male forensic psychiatric patients. Good inter-rater reliability was observed for the summary risk rating (for 27 women, ICC = .74; for 28 men, ICC = .69). In a different study, de Vogel et al. (2004) reported ICC_1 values for five combinations of rater groups (researchers; treatment supervisors; group leaders) on a sample of 53 male and 7 female forensic psychiatric patients. ICC values ranged from .63 (between treatment supervisors and group leaders) to .76 (between two researchers). Intermediate values included .65, .65, and .68. de Vogel and de Ruiter (2006) later reported that there were no significant differences in the ratings of low, moderate, or high risk between researchers, treatment supervisors, and clinical "group leaders" (responsible for day-to-day provision of services) who independently rated the cases.

In another study, Douglas, Yeomans, and Boer (2005) coded the HCR-20 on 188 male offenders released from prison. The IRR of the HCR-20 structured final judgment ratings, based on 28 cases, was fair to moderate (ICC = .41). Even with the lower IRR, there were no "category errors," or cases rated high risk by one rater and low risk by the other. Most of the disagreements were between moderate- and high-risk ratings. Finally, Douglas, Ogloff, and Hart (2003) coded the HCR-20 on a random sample of 100 forensic psychiatric patients who were released into the community. The HCR-20 was coded independently by two masters-level clinicians based on the

---

* One study of reliability reported percentage agreement instead of traditional reliability coefficients. In a sample of male intellectually disabled offenders, independent raters coded the HCR-20 on 30 cases using clinical files (Lindsay et al., 2008). The reliability was calculated based on the number of agreed-upon ratings divided by the number of agreements plus disagreements, expressed as a percentage. For the H subscale, agreement was 89.4%, for the C subscale 93.1%, and for the R subscale the agreement was 82.7%.

clinical-legal files of participants at the time of discharge from the hospital. Percentage agreement on the 50 IRR cases for the summary risk ratings was 70% with no instances of category errors (ICC = .61).

Inter-Rater Reliability by Setting.   As described above, the HCR-20 is used in a variety of settings,* and hence it is worthwhile to consider its reliability within those settings. Table 8.1 reports the median and range of reliability coefficients across different settings for the HCR-20 total and scale scores, along with the number of studies for each. Overall, across civil mental health, forensic mental health, corrections, and mixed samples, IRR for the HCR-20 total and H scale scores generally was excellent. There was more variability for the C and R scales, although IRR generally was strong for these subscales as well.

As shown in Table 8.1, the median IRR for the HCR-20 total score ranged from .79 (civil psychiatric) to .91 (corrections). For the H scale, it ranged from .83 (forensic psychiatric) to .94 (mixed samples). For the C scale, it ranged from .70 (civil psychiatric) to .83 (corrections). For the R scale, the median IRR ranged from .68 (mixed samples) to .90 (corrections). As mentioned above, there have been fewer studies of summary risk ratings. Reliabilities are generally lower than for the numeric scores; this is to be expected, given that it is a single item rating, rather than a multi-item scale.

Validity

In this section, we focus on the association between the HCR-20 and violence in terms of predictive validity. There have been other types of validity studied with the HCR-20, including its relationship with other risk assessment instruments, or its performance across different samples. However, the most important type of validity when it comes to risk assessment instruments is predictive validity. Hence, we will limit our discussion of validity research to this type of validity.

As with studies of the IRR of the HCR-20, the validity of this tool has been investigated extensively across a diversity of samples, settings, and countries. The HCR-20 was developed to assess risk for *general* violence, predominantly in populations where there is "a strong suggestion of mental illness or personality disorder" (Webster et al., 1997, p. 5). However, research on the HCR-20 has assessed its predictive validity with various violent outcomes including specific types of violence (e.g., sexual) and even nonviolent offenses (e.g., property crime, drug offenses). In a recent meta-analysis of 113 studies of the SPJ model, including 51 studies of the HCR-20 (Guy, 2008), the results indicated that effect sizes tended to be largest when SPJ instruments were used with the types of outcomes they were designed to be used with.

As with our review of the reliability of the HCR-20, we report the average validity effect sizes, rather than review dozens of studies individually. Study authors invariably use different types of effect sizes estimates, including correlation coefficients, odds ratios, and areas under the curve (AUCs) from receiver operating characteristic (ROC) analyses. Because the latter are most common in the risk assessment field, we converted other effect size estimates to AUCs for the purposes of this chapter. ROC analyses have been used commonly in the risk assessment field because they are less dependent on the base rate of the criterion variable in the sample (violence) than are other measures of association (e.g., correlation). AUC is an overall index of predictive

---

* Despite the fact that the HCR-20 was developed to assess general violence among adults, one study did investigate the use of the HCR-20 with adolescents. McEachern (2001) compared the HCR-20 and the SAVRY in a sample of 108 male juvenile offenders who were referred from court for inpatient psychiatric assessment (36 randomly selected nonrecidivists, 36 nonviolent recidivists, and 36 violent recidivists). This was a pseudo-prospective study conducted from comprehensive youth justice, police, mental health, medical, and social-demographic files. Inter-rater reliability, based on a subset of 36 files, was good (ICC_1 for Total, H, C, and R scores was .86, .88, .80, and .77, respectively).

Table 8.1 Median and Range of Inter-Rater Reliability Coefficients Across HCR-20 Studies

Setting	k*	HCR-20 Total Mdn [Range] (k)	H Scale Mdn [Range] (k)	C Scale Mdn [Range] (k)	R Scale Mdn [Range] (k)	SRR Mdn [Range] (k)
Forensic	21	.82 [.67–.95] (15)	.83 [.67–.94] (14)	.74 [.55–.94] (12)	.83 [.47–.94] (12)	.66 [.61–.76] (4)
Civil	3	.79 [.78–.80] (2)	.85 [.82–.87] (2)	.70 [.70] (1)	.81 [.81] (1)	—
Correctional	10	.91 [.67–.94] (7)	.90 [.58–.94] (5)	.83 [.58–.95] (6)	.90 [.58–.98] (5)	.41 [.41] (1)
Mixed	3	.90 [.90] (1)	.94 [.88–.97] (3)	.80 [.71–.89] (2)	.68 [.68] (1)	—

Studies include: Blum, 2004; Brown, 2004; Claix, Pham, & Willocq, 2002; Cooke, Michie, & Ryan, 2001; Côté, Hodgins, & Daigle, 2001; Dahle, 2006; Dernevik, 1998; Douglas, Ogloff, Nicholls, & Grant, 1999; Douglas, Ogloff, & Hart, 2003; Douglas, Yeomans, & Boer, 2005; Doyle & Dolan, 2006; Dunbar, 2003; Dunbar, Quinones, & Crevecoeur, 2005; Fujii, Tokioka, Lichton, & Hishinuma, 2005; Fujii, Lichton, & Tokioka, 2004; Gray, Snowden, MacCulloch, Phillips, Taylor, & MacCulloch, 2004; Gray, Fitzgerald, Taylor, MacCulloch, & Snowden, 2007; Grevatt, Thomas-Peter, & Hughes, 2004; Hildebrand, Hesper, Spreen, & Nijman, 2005; Hill, Habermann, Klusmann, Berner, & Briken, 2008; Howard, 2007; Kroner & Mills, 2001; MacPherson & Kevan, 2004; McDermott, Edens, Quanbeck, Busse & Scott, 2008; McNiel, Gregory, Lam, Binder, & Sullivan, 2003; Mills, Kroner, & Hemmati, 2007; Müller-Isberner, Sommer, Özokyay, & Freese, 1999; Pham, Claix, & Remy, 2000; Philipse, 2002; Polvi, 1999; Ross, Hart, & Webster, 1998; Strand, Belfrage, Fransson, & Levander, 1999; Warren, South, Burnette, Rogers, Friend, Bale, & Van Patten, 2005; Wintrup, 1996.

* Number of studies refers to the total number of studies within a given setting, rather than the number of studies that reported IRR data. Mixed samples include studies that combined data from more than one type of sample (correctional, civil mental health, forensic mental health). SRR = summary risk rating (low, moderate, high). For cells in which there is only one effect size summarized, that effect size also is reported as the median and range. *k* = number of studies.

accuracy, and can range from 0 (perfect negative prediction), to .50 (chance prediction), to 1.0 (perfect positive prediction). A given area represents the probability that a randomly chosen person who scores positive on the dependent measure (i.e., is actually violent) will fall above any given cutoff on the predictor measure, and that an actually nonviolent person will score below the cutoff (Mossman & Somoza, 1991). Another interpretation is that the AUC represents the probability that a violent person would score higher than a nonviolent person on the predictor. Although there are no formal categories, AUC values of approximately .65 to .70 may be considered moderate to large, and approximately .70 and above may be considered large (see, e.g., Douglas et al., 2005; Rice & Harris, 2005).

Overall, the results of predictive validity studies show that the association between the HCR-20 and violence is of moderate to large magnitude. The median AUC value for the HCR-20 total score across 42 studies is .69. This effect size is comparable to that for actuarial instruments designed to assess violence risk (.67) across 45 studies (Guy, 2008). Further, the effect size for the HCR-20 includes all studies, regardless of the type of outcome. As mentioned above, some studies have focused on the HCR-20's ability to predict specific types of violence (for example, sexual

Table 8.2 The Range of AUC Values for the HCR-20 Total, Subscales, and Summary Risk Rating

AUC	HCR-20 Total ($k = 42$)	H ($k = 38$)	C ($k = 33$)	R ($k = 29$)	SRR ($k = 6$)
.30–.39	1	1	1	0	0
.40–.49	2	3	5	3	0
.50–.59	14	14	18	12	0
.60–.69	36	34	29	27	3
.70–.79	29	25	20	18	5
.80–.89	15	6	1	2	2*
Total	97	83	74	62	10

Studies include: Allen & Howells, 2008; Cooke, Michie, & Ryan, 2001; de Vogel & de Ruiter, 2005; de Vogel & de Ruiter, 2006; de Vogel, de Ruiter, Hildebrand, Bos, & van de Ven, 2004; Dernevik, Grann, & Johansson, 2002; Dolan & Fullam, 2007; Dolan & Khawaja, 2004; Douglas, Ogloff, & Hart, 2003; Douglas, Ogloff, Nicholls, & Grant, 1999; Douglas, Yeomans, & Boer, 2005; Doyle & Dolan, 2006; Doyle, Dolan, & McGovern, 2002; Fujii, Tokioka, Lichton, & Hishinuma, 2005; Fujii, Lichton, & Tokioka, 2004; Grann, Belfrage, & Tengstrom, 2000; Grann, Sturidsson, Haggard-Grann, Hiscoke, Alm, & Dernevik, et al, 2005; Grann & Wedin, 2002; Gray, Hill, McGleish, Timmons, MacCulloch, & Snowden, 2003; Gray, Fitzgerald, Taylor, MacCulloch, & Snowden, 2007; Gray, Snowden, MacCulloch, Phillips, Taylor, & MacCulloch, 2004; Gray, Taylor, & Snowden, 2008; Grevatt, Thomas-Peter, & Hughes, 2004; Hartvig, Alfarnes, Skjonberg, Moger, & Ostberg, 2006; Kroner, & Mills, 2001; Lindsay, Hogue, Taylor, Steptoe, Mooney, O'Brien, et al., 2008; MacPherson, & Kevan, 2004; McDermott, Edens, Quanbeck, Busse, & Scott, 2008; McKenzie, & Curr, 2005; McNiel, Gregory, Lam, Binder, & Sullivan, 2003; Mills, Kroner, & Hemmati, 2007; Morrissey, Hogue, Mooney, Allen, Johnston, Hollin, Lindsay, & Taylor, 2007; Neves, & Gonçalves, 2008; Nicholls, 2001; Nicholls, Ogloff, & Douglas, 2004; Pham, Ducro, Marghem, & Réveillère, 2005; Philipse, 2002; Polvi, 1999; Ross, Hart, & Webster, 1998; Stadtland, & Nedopil, 2005; Stadtland, Hollweg, Kleindienst, Dietl, Reich, & Nedopil, 2005; Strand, Belfrage, Fransson, & Levander, 1999; Tengström, 2001; Tengstrom, Hodgins, Muller-Isberner, Jockel, Freese, & Ozokyay, et al., 2006; Urheim, Jakobsen, & Rasmussen, 2003; Warren, South, Burnette, Rogers, Friend, Bale, & Van Patten, 2005.

Note: k = number of studies. Values in the "Total" row are number of effect sizes, not number of studies.
* One AUC of .91.

assault) or nonviolent antisocial behavior. Guy (2008) reported that, when studies included outcomes that more closely conformed to the HCR-20 definition of violence, effect sizes tended to be higher. For example, studies that used the HCR-20 definition of violence, on average, produced a slightly higher weighted mean AUC of .73.

The median effect sizes for the H, C, and R scales are somewhat lower than for its total score, at .68, .62 and .65, respectively. Table 8.2 provides a breakdown of effect sizes as a function of scale and effect size range. The AUCs for the H scale ranged from .40 (Grevatt, Thomas-Peter, & Hughes, 2004) to .85 (Nicholls, Ogloff, & Douglas, 2004). For the C scale, the AUCs ranged from .46 (Nicholls et al., 2004) to .80 (de Vogel & de Ruiter, 2006). Finally, the range for R scale AUCs was from .48 (Nicholls, 2001) to .88 (de Vogel & de Ruiter, 2005). The AUCs for the HCR-20 have been broken down by setting as well (see Table 8.3 through Table 8.6).

Overall, the research indicates that the HCR-20 total and scale scores are associated with violence, with average effect sizes of moderate to moderate/large magnitude, comparable to actuarial risk assessment measures. This establishes that the risk factors, as defined, are associated with violence and, hence, ostensibly could serve as a basis for evaluators to make decisions about risk, a topic we turn to next.

Validity of Summary Risk Ratings.   As described above, the summary risk ratings of low, moderate, or high risk are commonly used in the SPJ model to communicate concerns about

SRR
(k = 6)
0
0
0
3
5
2*
10

05; de Vogel
k, Grann, &
Hart, 2003;
Dolan, 2006;
Lichton, &
n, Hiscoke,
MacCulloch,
, Snowden,
08; Grevatt,
; Kroner, &
n, & Kevan,
05; McNiel,
ey, Hogue,
holls, 2001;
ipse, 2002;
leindienst,
röm, 2001;
akobsen, &

studies.

uded out-
tended to
age, pro-

tal score,
nction of
-Peter, &
ged from
le AUCs
-20 have

with vio-
to actu-
sociated
ns about

 of low,
is about

Table 8.3 The Range of AUC Values in Forensic Psychiatric Samples for the HCR-20 Total, Subscales, and Summary Risk Rating

AUCs	HCR-20 Total (k = 27)	H (k = 24)	C (k = 23)	R (k = 19)	SRR (k = 6)
.30–.39	1	1	0	0	0
.40–.49	0	1	4	1	0
.50–.59	9	8	9	10	0
.60–.69	20	18	15	19	2
.70–.79	16	8	12	4	3
.80–.89	8	4	1	1	2*
Total	54	40	41	35	7

Studies include: Allen & Howells, 2008; de Vogel & de Ruiter, 2005; de Vogel & de Ruiter, 2006; de Vogel, de Ruiter, Hildebrand, Bos, & van de Ven, 2004; Dernevik, Grann, & Johansson, 2002; Dolan & Fullam, 2007; Dolan & Khawaja, 2004; Douglas, Ogloff, & Hart, 2003; Fujii, Tokioka, Lichton, & Hishinuma, 2005; Fujii, Lichton, & Tokioka, 2004; Grann, Belfrage, & Tengstrom, 2000; Grann, Sturidsson, Haggard-Grann, Hiscoke, Alm, & Dernevik, et al, 2005; Gray, Hill, McGleish, Timmons, MacCulloch, & Snowden, 2003; Gray, Snowden, MacCulloch, Phillips, Taylor, & MacCulloch, 2004; Gray, Taylor, & Snowden, 2008; Gray, Fitzgerald, Taylor, MacCulloch, & Snowden, 2007; Grevatt, Thomas-Peter, & Hughes, 2004; Lindsay, Hogue, Taylor, Steptoe, Mooney, O'Brien, et al., 2008; MacPherson, & Kevan, 2004; McDermott, Edens, Quanbeck, Busse, & Scott, 2008; McKenzie, & Curr, 2005; Morrissey, Hogue, Mooney, Allen, Johnston, Hollin, Lindsay, & Taylor, 2007; Nicholls, 2001; Philipse, 2002; Polvi, 1999; Stadtland, & Nedopil, 2005; Strand, Belfrage, Fransson, & Levander, 1999; Tengström, 2001; Tengstrom, Hodgins, Muller-Isberner, Jockel, Freese, & Ozokyay, et al., 2006; Urheim, Jakobsen, & Rasmussen, 2003.

Note: k = number of studies. Values in the "Total" row are number of effect sizes, not number of studies.
* One AUC of .91.

Table 8.4 The Range of AUC Values in Civil Psychiatric Samples for the HCR-20 Total, Subscales, and Summary Risk Rating

AUCs	HCR-20 Total (k = 5)	H (k = 4)	C (k = 4)	R (k = 4)	SRR
.30–.39	0	0	0	0	0
.40–.49	0	0	1	0	0
.50–.59	1	4	5	1	0
.60–.69	9	11	13	1	0
.70–.79	11	9	1	12	0
.80–.89	4	2	0	0	0
Total	25	26	20	14	0

Studies include: Douglas, Ogloff, Nicholls, & Grant, 1999; Hartvig, Alfarnes, Skjonberg, Moger, & Ostberg, 2006; McNiel, Gregory, Lam, Binder, & Sullivan, 2003; Nicholls, Ogloff, & Douglas, 2004; Ross, Hart, & Webster, 1998.

Note: k = number of studies. Values in the "Total" row are number of effect sizes, not number of studies.

level of risk and degree of intervention required to mitigate that risk. This aspect of the HCR-20, and other SPJ tools, has been investigated less commonly than the numeric use of the instruments. Because this risk estimate is nonactuarial, it is perceived as controversial and criticized by some commentators (Quinsey et al., 2006; Rice et al., Chapter 6, this volume). Across all SPJ tools, however, there have been 16 studies of summary risk ratings, and evidence indicates that they are as or more strongly related to violence than is the numeric use of the instruments, or the use of actuarial instruments to which they have been compared, including the Static-99, the VRAG, and the SORAG (Guy, 2008; Heilbrun et al., 2009).

Table 8.5 The Range of AUC Values in Correctional Samples for the HCR-20 Total, Subscales, and Summary Risk Rating

AUCs	HCR-20 Total (k = 6)	H (k = 3)	C (k = 2)	R (k = 2)	SRR (k = 1)
.30–.39	0	0	0	0	0
.40–.49	0	0	0	0	0
.50–.59	2	0	0	0	0
.60–.69	5	4	0	0	0
.70–.79	3	2	2	1	1
.80–.89	3	0	0	1	0
Total	13	6	2	2	1

Studies include: Cooke, Michie, & Ryan, 2001; Douglas, Yeomans, & Boer, 2005; Kroner, & Mills, 2001; Mills, Kroner, & Hemmati, 2007; Neves, & Gonçalves, 2008; Warren, South, Burnette, Rogers, Friend, Bale, & Van Patten, 2005.

Note: k = number of studies. Values in the "Total" row are number of effect sizes, not number of studies.

Table 8.6 The Range of AUC Values in Mixed Samples for the HCR-20 Total, Subscales, and Summary Risk Rating

AUC	HCR-20 Total (k = 3)	H k = 5)	C (k = 2)	R (k = 2)	SRR
.30–.39	0	0	0	0	0
.40–.49	1	0	0	1	0
.50–.59	0	0	1	0	0
.60–.69	3	3	1	2	0
.70–.79	3	3	1	0	0
.80–.89	0	0	0	0	0
Total	7	6	3	3	0

Studies include: Doyle & Dolan, 2006; Doyle, Dolan, & McGovern, 2002; Grann & Wedin, 2002; Pham, Ducro, Marghem, & Réveillère, 2005; Stadtland, Hollweg, Kleindienst, Dietl, Reich, & Nedopil, 2005.

Note: k = number of studies. Values in the "Total" row are number of effect sizes, not number of studies.

Six studies have investigated the predictive accuracy of the HCR-20's summary risk ratings. In the first such study, Douglas et al. (2003) completed the HCR-20 for 100 randomly selected forensic psychiatric patients using clinical–legal files as they existed at the time of the patients' discharge to the community. Violent recidivism, detected through multiple file-based sources, was divided into three categories: any violence, physical violence, and non-physical violence. The HCR-20 summary risk ratings were statistically significant and of moderate to large size (any violence, AUC = .69; physical violence, AUC = .74; nonphysical violence, AUC = .68).

de Vogel and de Ruiter (2005; as described in detail above) coded the HCR-20 using file information of Dutch forensic psychiatric patients. For females, 15 cases were coded retrospectively (with inpatient violence as the criterion) and 27 cases prospectively (with postdischarge community violence as the criterion). For males, half were coded retrospectively, and half were coded prospectively (again, with inpatient and community violence as the respective criteria). The averaged predictive validity of the summary risk ratings across both community and inpatient violence were large for both men (AUC = .91) and women (AUC = .86).

In a second study of Dutch forensic psychiatric patients by de Vogel and de Ruiter (2006), the predictive validity of the HCR-20 was investigated prospectively in a sample of 127 males. Three groups of coders were compared: researchers, treatment supervisors, and group leaders.

Historical-Clinical-Risk Management-20 (HCR-20) Violence Risk Assessment Scheme • **169**

In addition, a consensus score was obtained—after independent ratings were made, evaluators discussed each case for approximately one hour in order to reach agreement about the HCR-20 item scores, as well as the summary risk rating of low, moderate, or high risk. The outcome measure was again both inpatient and community violence recorded after a mean follow-up period of 21.5 months for inpatients and 15 months for those discharged into the community. The AUCs for violence for the HCR-20 summary risk ratings for the three groups of raters ranged from .64 to .77. However, the consensus rating was substantially higher, at .86. This finding of a larger AUC for consensus ratings provides some indication that, if possible, team-based risk assessments might provide more accurate risk estimates than individual risk ratings. The AUC for both verbal abuse (consensus AUC = .65) and verbal threat (consensus AUC = .71) were significant as well. In a third study by de Vogel and colleagues (2004), 120 forensic patients who were discharged from a forensic facility were followed in the community for an average of 73 months. The AUC for the SPJ judgment was .79.

Fujii, Lichton, and Tokioka (2004) studied the HCR-20 summary risk ratings in a sample of 169 forensic patients. The HCR-20 was administered during the first week of the patients' hospital admission. Episodes of inpatient violence (as defined in the HCR-20 manual) were recorded from hospital event records for a minimum of three months postadmission. The final risk judgment (in this study, high risk versus low/moderate risk) produced a significant AUC of .70.

Finally, HCR-20 summary risk ratings have been investigated in a correctional sample of 188 male offenders released from prison and followed for 6 to 11 years (Douglas et al., 2005). The HCR-20 was coded from extensive file information as it existed at the time of release. This study used a known-groups design, with 93 recidivistic participants being matched to 95 nonrecidivists. Reconviction and reimprisonment for a violent offense was the outcome variable, coded blind to HCR-20 ratings or group status. The AUC for the HCR-20 SPJ final judgment was large and significant (AUC = .79).

As with HCR-20 numeric scores, effect sizes for the summary risk rating tend to be higher when studies use the definition of violence contained in the HCR-20, rather than specialized forms of violence, or nonviolent antisocial behavior. Guy (2008) reported that the average AUC for the HCR-20 summary risk rating, regardless of type of outcome, was .70. However, when studies used the HCR-20 definition of violence, the average AUC was .76. For analyses involving physical violence only, it was .79.

It is possible to report the proportion of persons who are violent in any given sample as a function of estimated risk level. We have done so in Table 8.7. We caution that these percentages are not intended to represent numeric probability estimates that should be used in clinical practice. Rather, they are presented to indicate that, if evaluators choose to use the HCR-20 and its summary risk rating system, there is evidence that their judgments have some meaning.

Comparison of Summary Risk Ratings to Actuarial Ratings of Risk. The most basic test of the comparative validity of the summary risk ratings is a simple bivariate comparison between their predictive validity and the predictive validity of the HCR-20 numeric scores. If evaluators are using summary risk ratings as opposed to numeric scores to make their risk estimates, it is important that the summary risk ratings are at least as accurate as the numeric use of the HCR-20, if not more so. All six studies of the HCR-20 summary risk ratings permit this simple comparison. Comparing across nine sets of effect sizes from these six studies, seven favored the summary risk rating over the HCR-20 total score, in terms of the size of the AUC. For the two comparisons where the HCR-20 total score produced an AUC that was greater than that for the HCR-20 summary risk rating, it was greater by .03 and .04. For the seven comparisons that produced a larger AUC for the summary risk ratings compared to the HCR-20 total score, the differences tended to be small (.01., .01, .02, .03, .04), although two were more substantial (.09, .27).

Table 8.7 Percent Violent as a Function of HCR-20 Low, Moderate, and High Risk Ratings

Study	N	Setting	Risk Category		
			Low % (n_v/n_t)	Moderate % (n_v/n_t)	High % (n_v/n_t)
de Vogel et al. (2004)	119	Forensic	0 (0/14)	15 (7/47)	62 (36/58)
de Vogel & de Ruiter (2005)					
Men	42	Forensic	0 (0/11)	8 (1/13)	78 (14/18)
Women	41	Forensic	0 (0/11)	14 (3/21)	77 (7/9)
de Vogel & de Ruiter (2006)	127	Forensic	0 (0/36)	8 (5/61)	64 (19/30)
Douglas et al. (2003)	100	Forensic	4 (1/23)	11 (7/64)	54 (7/13)
Douglas et al. (2005)	188	Corrections	19 (13/68)	59 (50/85)	86 (30/35)
Fujii et al. (2004)	169	Forensic		17 (22/127)	55 (18/33)

Note: n_v = number of participants who were violent within the estimated risk categories of low, moderate, or high risk; n_t = number of participants estimated to be low, moderate, or high risk; Fujii et al. (2004) collapsed the low and moderate risk categories for their analyses, which we report here. Values for de Vogel et al. (2004) calculated from raw data provided by Vivienne de Vogel.

On average, the AUC for the summary risk rating was greater than that for the HCR-20 total score by .064. The AUC for the summary risk rating was substantially larger than the HCR-20 H, C, and R scale scores in these studies, as these indices tend to be less strongly related to violence than the HCR-20 total score.

A more meaningful test of the comparative validity of the HCR-20 summary risk ratings and the HCR-20 numeric scores comes from multivariate rather than bivariate comparisons. Douglas et al. (2003) were the first study to report such an analysis, in the study described above of 100 forensic psychiatric patients. In Cox regression analyses, the H, C, and R scales were entered as the first block of predictors, with physical violence in the community as the outcome, and produced a significant model fit, $\chi^2 = 9.9$, $p < .05$. On the second block of predictors, the HCR-20 summary risk rating of low, moderate, and high risk was entered, producing a significant model improvement ($\Delta\chi^2 = 9.8$, $p < .01$) and final model ($\chi^2 = 20.07$, $p < .0001$). Importantly, *only* the summary risk ratings were significant in this final model, with an associated hazard ratio of 9.4, indicating that they increased the hazard of violence ninefold at each step of the predictor (i.e., from low to moderate risk, and from moderate to high risk).

de Vogel and de Ruiter (2006) reported the outcome of a similar analysis in their prospective study of 127 male Dutch forensic patients. They used Cox regression analyses with the HCR-20 subscales entered on the first block and the summary risk rating entered on the second block to determine if the final risk judgment added incremental validity to the subscale scores. Block one produced a significant model (χ^2 [3, $N = 127$] = 22.9, $p < .001$), and the HCR-20 summary risk ratings demonstrated incremental validity on block 2 as there was significant improvement to the model's fit upon their entry, $\Delta\chi^2$ (1, $N = 127$) = 6.8, $p < .01$. de Vogel et al. (2004) reported comparable findings from their other study of Dutch forensic patients.

Fujii et al. (2004) took a somewhat different approach, expanding the summary risk rating to five levels (low, low–moderate, moderate, moderate–high, and high). A stepwise regression demonstrated that the five-level summary risk rating added incremental validity over HCR-20 numeric scores (an increase in R^2 from .036 to .092), whereas the reverse was not true.

Finally, the HCR-20 summary risk ratings have been tested against *other* decision-making approaches, including unstructured clinical prediction, the PCL-R, PCL:SV, VRAG, and an actuarial instrument called the VORAS (Violent Offender Risk Assessment Scale) (Howells, Watt, Hall, & Baldwin, 1997). De Vogel et al. (2004), for instance, tested the summary risk rating against the HCR-20 numeric score, the PCL-R, as well as unstructured predictions of violence. Using Cox regression, they first entered the index of unstructured prediction, which produced a significant model fit. On the second block of predictors, they entered the PCL-R and HCR-20 total scores. This block also improved predictive accuracy, with the HCR-20 total scores accounting for the improvement to model fit. Finally, on block 3, they entered the HCR-20 summary risk ratings, which again produced a significant improvement to model fit, with both the summary risk ratings and the HCR-20 numeric score being significant individual predictors in the final model. The effect size (hazard ratio) for the summary risk ratings was 3.1, indicating that the hazard for violence tripled at each step of the predictor (from low to moderate risk, and from moderate to high risk).

Douglas et al. (2005) conducted a set of analyses using binary logistic regression that included the HCR-20 summary risk ratings, the VRAG's actuarial categorical system, and the VORAS. All variables were significant predictors of violent recidivism. However, the VORAS performed poorly, being inversely related to violence. As such, a second logistic regression model was tested that included only the HCR-20 summary risk ratings and the VRAG nine-level categorical system—the two indices intended to be used in practice. In the final model, both were significant predictors (odds ratio for the HCR-20 = 2.90, for the VRAG, 1.63). This result indicates that the nonactuarial HCR-20 summary risk rating and the actuarial VRAG risk estimate each offered unique predictive power.

Although only six studies have investigated the HCR-20 summary risk ratings, results support their utility, and are consistent with the 10 or so other studies of summary risk ratings of other SPJ instruments (Guy, 2008). In summary, all four studies that *directly* have tested the summary risk rating against the numeric use of the HCR-20 in multivariate analyses have shown that it adds incrementally to the numeric use of the instrument, often to the extent that the HCR-20 numeric score is no longer significant in the predictive model. Studies that have tested the summary risk ratings against other decision-making approaches (PCL-R; VORAS; VRAG; unstructured clinical prediction) have shown that it either is more strongly related to outcome than those other instruments, adds incrementally to them, or possesses unique predictive variance in the presence of them. This line of research is important, because the summary risk ratings are intended to be the basic method of communicating an evaluator's judgment about level of risk and anticipated degree of intervention. Further, the summary risk rating is nonactuarial and has been criticized by some on that basis (Quinsey et al., 2006; Rice et al., this volume). However, across six HCR-20 studies and ten other SPJ studies, results are consistent that summary risk ratings are as or more strongly related to violence than are numeric estimates of risk produced by SPJ instruments, or by other decision-making procedures, including actuarial ones.

Limitations and Necessary Future Research

Despite more than 50 empirical evaluations of the HCR-20, it has limitations, as do all risk assessment instruments. First we outline some of these limitations, and then discuss fruitful avenues for future research.

Limitations

Format of Risk Estimates.   According to some commentators, the fact that the HCR-20 does not provide numeric estimates of future risk for violence (at least for clinical purposes) is a limitation of the HCR-20, and all SPJ instruments (Quinsey et al., 2006; see also Heilbrun et al., Chapter 1, this volume). We mention this criticism of the HCR-20 to acknowledge that it is considered a limitation by some. However, at the same time, we do not actually consider this a true limitation of the HCR-20, but rather a choice that we believe best reflects the state of the discipline. As discussed in detail above, although instruments that produce numeric estimates of risk for violence have the appearance of precision when it comes to estimating future risk for violence, research has to date failed to support the generalizability of such "precise" estimates upon cross-validation. Therefore, in our view a nonnumeric risk estimation system avoids this pitfall while, at the same time, provides guidance about relative risk, relevant risk factors, and the degree and nature of intervention, supervision, monitoring, and management required to mitigate risk.

Gender and Ethnicity.   As with all existing risk assessment instruments, we do not yet know enough about the potential influence of gender and ethnicity on the utility of the HCR-20. Some have argued, for instance, that certain risk factors might be more important for women than men (Holtfreter & Cupp, 2007). Further, research indicates that some risk factors (i.e., psychopathy, as measured by the PCL-R), may manifest somewhat differently for people of different ethnic or racial backgrounds (Cooke, Kosson, & Michie, 2001).

Although we do consider further research on gender and ethnicity important and necessary, in our view the HCR-20 likely is robust across such demographic groups, for several reasons. First, because the HCR-20 items were neither selected from nor optimized statistically on a single sample, but rather chosen based on the broader literature, the potential for any of the risk factors to be dependent on *any* sample characteristic is lower than for instruments that did select and optimize risk factors and predictive algorithms based on single samples.

Second, research indicates that the HCR-20 performs comparably across genders and countries. The few studies of the HCR-20 with women do not provide any evidence that its predictive validity is lower than for men (de Vogel & de Ruiter, 2006; Nicholls et al., 2004; Strand & Belfrage, 2001). One study did indicate poor performance with a sample of women (Warren et al., 2005). However, this was a postdictive study in which the HCR-20 was used to "predict" past convictions of murder. Hence, it is of little relevance to the predictive validity of the HCR-20. Similarly, the HCR-20 has been evaluated in approximately 12 countries, with no discernible difference in predictive utility. In fact, in a meta-analysis of the larger SPJ literature (113 studies), Guy (2008) reported that neither gender nor country moderated the relationship between SPJ instruments and violence. Indeed, there was a nonsignificant ($p = .06$) trend for instruments to perform better for women than for men.

Because the HCR-20 (both its numeric scores and summary risk ratings made by researchers or clinicians) predicts violence at comparable levels across numerous countries, and because countries differ in terms of history, customs, sociopolitical leanings, ethnic mix, and laws, there is strong evidence for the robustness of the HCR-20 across ethnicities. However, there has been less research on its utility across ethnic groups *within* countries. Because the HCR-20 risk factors were neither selected nor optimized on any particular sample, however, there is reason to believe that it would be robust across different ethnic groups within countries. Fujii, Tokioka, Lichton, and Hishinuma (2005), in a sample of 169 forensic psychiatric patients, compared the frequency of HCR-20 risk factors and its predictive validity across 51 Asian American patients, 46 Euro-American patients, and 38 Native Hawaiian patients. Some risk factors were more

prevalent than others across groups, and there were some differences in which risk factors were most predictive across groups. Although there were no significant differences in the predictive utility of the HCR-20 Total scores across groups, its performance was best among Native Hawaiians (AUC = .73) and worst among Asian Americans (AUC = .58). Because there was not a significant moderating effect for ethnicity, Fujii et al. concluded that their data indicated that the "HCR-20 has cross-cultural validity in Asian-American, Native-Hawaiian, and Euro-American samples" (p. 714).

We have two comments about the line of research pursued by Fujii et al. (2005). First, generally, it is an important line of research that can address whether the HCR-20, or any risk assessment instrument, performs comparably across different ethnic groups. Second, we would caution researchers to be somewhat wary about item-level tests that are based on single samples, because they are likely to be highly sample-specific. In fact, this is precisely the problem faced by actuarial instruments that select or weight single items based on single samples. There is no guarantee that any differences observed across ethnic (or any) groups at the item level are robust, and not due to sample peculiarities. For this reason, any differences in item-level frequency, reliability, or predictive validity should be (1) tested in large subgroups and (2) observed across multiple independent samples before firm conclusions can be drawn with confidence about reliable differences.

Nonetheless, we encourage evaluators to attend to any research on how certain risk factors may manifest differently across ethnic groups, particularly if there is research indicating differential predictive validity of such risk factors. In our view, the larger research on violence indicates that most risk factors (i.e., previous violence; substance use problems) are important regardless of ethnicity.

Similarly, we encourage evaluators to attend to how risk factors might manifest differently for men and women. Although we doubt that most risk factors will be relevant for one gender but not for the other, it is possible that certain risk factors might be more prevalent in one group versus another (i.e., childhood sexual abuse). It is also possible that there are "gendered" pathways to violence (Holtfreter & Cupp, 2007), and that the methods and expression of violence might vary across genders.

Because of the way that the HCR-20 is intended to be used, however, we consider such possible different manifestations of violence and violence risk factors across genders or ethnicities to be within the scope of the intended use of the HCR-20. That is, risk factors are unit-weighted, and evaluators are expected to describe both the individual manifestation of risk factors and their relevance for understanding violence, for the case at hand (that is, what does the risk factor *look like,* for this person, and how is it relevant to his or her risk for violence?).

Future Research

We encourage more research on all of the topics discussed under "Limitations," above. In addition, there are several other areas of research that could prove to be fruitful, both for further understanding of the HCR-20, and also for understanding about risk assessment in general.

Understanding the Effectiveness of Nonnumeric Categorical Risk Ratings. As reviewed above, research on the HCR-20 (and other SPJ instruments) indicates that its system of nonnumeric, categorical summary risk ratings performs as well as or better than the numeric (actuarial) use of the instrument, the PCL-R, and actuarial instruments such as the VRAG (and, in tests of other SPJ instruments, the Static-99). This finding has been robust across more than a dozen tests, across countries, and whether the research was conducted by individuals associated with the HCR-20 (or other SPJ instruments) or not. This finding is also potentially controversial, in that the categorical summary risk ratings, while based on a structured system, are not actuarial,

but discretionary. The question arises, *why* do these discretionary judgments perform as well as or better than numeric/actuarial estimates? There is not a single study on point. We offer four possibilities that would make for interesting research studies.

First, it could be that the SPJ method of categorical risk ratings strikes an optimal balance between structure and discretion. Although structure is necessary to promote good decision making, it is possible that current actuarial risk assessment measures are overstructured. That is, they are fine-tuned as a function of development samples and may not fit as well when applied in different samples. Further, as part of their structure, most presume that all risk factors operate equally for all persons, rather than permitting a determination of which risk factors may be more or less important in individual cases. The SPJ model facilitates this latter aspect of decision making, and it could be that this relatively limited degree of discretion, couched within an otherwise structured decision-making process, improves accuracy.

Second, and related to applying discretion at the individual level, SPJ measures such as the HCR-20 may represent an effective method of bridging the nomothetic and idiographic levels of analysis. That is, SPJ measures include a standard, fixed list of risk factors that must be considered in every case. These factors have support in the broad empirical literature. However, there is not a presumption that each of these nomothetically supported risk factors apply equally at the idiographic or case level. Indeed, as with many phenomena in nature, the individual relevance of risk factors likely is normally distributed. Most statistical approaches that are used to develop or validate actuarial measures, however, are premised on group-based estimates (i.e., beta coefficients) that essentially average across all persons in a sample to determine the predictive strength of a variable for a *sample*, not for any given *person* within that sample. The SPJ approach embodied by the HCR-20, however, treats the scoring of nomothetically supported risk factors as a starting point in an individually based clinical assessment process. Additional aspects of this assessment process include explicating the individual manifestation and relevance of such risk factors for the given case at hand. In essence, this task is akin to determining where, in the distribution of a risk factor's potential relevance to violence, a person falls. This step is not represented in any actuarial measures that we are aware of, and may serve to optimize the relevance of nomothetic data at the idiographic level.

Third, the SPJ process may facilitate a task that even Meehl (1954)—the strongest, most persuasive, and most vocal advocate of the superiority of actuarial prediction over clinical prediction—conceded was likely better accomplished by people rather than algorithms: derivation of theory, and the recognition of configural relations or patterns. Many actuarial risk assessment instruments presume relatively simple relationships between risk factors and violence (a main-effect relationship where an increase in a risk factor elevates the odds of violence). Even actuarial instruments based on more complex statistical processes (that is, interactions between variables) ultimately presume a relatively simple relationship between risk factors and violence, and they do so, again, with the assumption that all risk factors in a predictive equation should be weighted equally for all persons.

The SPJ process encourages decision makers to build "individual theories" of violence for each person they evaluate. It may facilitate the identification of "configural relations" between a set of risk factors and violence, one in which risk factors might not only interact with one another, but may transact with one another, and with violence. It encourages decision makers to search for patterns amongst risk factors that may be of particular relevance for whether a person will be violent. Importantly, while guided by the nomothetic literature that essentially informs us that "the more risk factors present, the greater the risk," it does not presume that risk-relevant patterns or relations between risk factors will be uniform across individuals. In this way, it may be better suited to help clinicians *understand* the violence of individuals, rather than merely to predict it. It is geared toward identifying the most relevant, potentially causal, risk factors, and

patterns among risk factors, *at the individual level*. This aspect of the SPJ model could be another reason that explains why the relatively basic judgments of low, moderate, and high risk are as or more accurate in forecasting future violence compared to other approaches.

Finally, it is possible that the risk judgments of low, moderate, or high risk simply permit the consideration of additional information above and beyond what is captured by the standard list of risk factors on SPJ instruments, or the actuarial instruments to which they sometimes have been compared. Indeed, SPJ measures encourage evaluators to consider whether any other risk-relevant information exists for any given case. Despite attempts to provide relatively comprehensive domain coverage of risk factors, it is always possible that there may be risk factors present for a given case that are not well represented amongst the standard risk factors contained in the instrument.

It has long been determined that a "fair test" between clinical and actuarial prediction can only be had if the two methods of prediction are based on the same information (Grove et al., 2000). We agree, but demur. The HCR-20 was never developed to inform the classic "actuarial versus clinical prediction" debate. Rather, it was developed to promote sound (reliable and valid) decisions about violence risk that also are relevant to risk reduction at the case level. At the same time, our goal was to ensure that it was no *less* valid than actuarial procedures, which it does not appear to be. As such, if it is effective because it permits the consideration of additional relevant information, so be it.

Reducing Violence. Ultimately, the point of assessing the risk for violence is to prevent future violence (Douglas & Kropp, 2002; Douglas & Skeem, 2005; Hart, 1998). The HCR-20 was developed in the manner it was in order to inform risk management and, hence, to reduce a person's risk for violence. However, this aspect of the HCR-20 has not yet received the research attention that it needs. That is, by using the HCR-20 to identify a person's key risk factors and to design optimal risk management plans, will the person be less likely to be violent in the future than if the HCR-20 were not used? To answer this question requires moving from risk assessment to intervention research, which we would encourage with enthusiasm.

There is indirect evidence to support the hypothesis that the appropriate use of the HCR-20 would reduce violence. Research based on a short version of the SARA (Spousal Assault Risk Assessment, an SPJ measure for spousal violence; see Chapter 11, this volume) used with police officers showed that, when officers rated cases as high risk, and also applied a commensurate degree of intervention (that is, a high level of appropriate interventions), future spousal violence was reduced in comparison to high-risk cases in which a commensurate degree of intervention was not enacted.

Version 3 of the HCR-20. The HCR-20 currently is under revision, and will be named HCR:V3 (Douglas, Hart, Webster, Belfrage, & Eaves, 2008). We are in the process of conducting beta testing, user feedback, and studies of its reliability and validity. The fundamental SPJ approach will be retained in HCR:V3, although there may be changes to items. New scoring options are being tested, as well. Evaluation of HCR:V3 will be an important research endeavor.

Case Example

Mr. Case is a 53-year-old Caucasian male who was first admitted to the Secure Forensic Hospital (SFH) on June 14, 1982, for a competency evaluation after he was charged with second degree murder. He was found competent to stand trial, but was readmitted to SFH on October 9, 1982, when he was found Not Guilty by Reason of Insanity (NGRI) for the same charge. Mr. Case has a Review Board hearing scheduled for next month

to determine whether he should (1) be discharged without conditions; (2) discharged with conditions; or (3) not discharged.

A comprehensive violence risk assessment was conducted to inform the decision of the Review Board. The key questions addressed were: (1) What is the level of risk posed by Mr. Case, if he is released into the community? (2) What risk management strategies would mitigate this risk? The risk assessment included a lengthy interview with Mr. Case and a review of his institutional files. Given Mr. Case's lengthy forensic hospitalization, it was not considered necessary to conduct additional psychological testing. The Psychopathy Checklist–Revised (PCL-R) had recently been completed for Mr. Case by a psychologist with training on the instrument and was used to score HCR-20 Item H7 (Psychopathy) for the present assessment.

FAMILY HISTORY

Mr. Case was raised in Ohio by his natural parents and has one older brother. He had a good relationship with his parents and brother when younger, but fought frequently both verbally and physically with his brother starting at age 16. Mr. Case claims that the fights with his brother usually started as minor "sibling disagreements," but would quickly escalate into physical fights. He stated that both he and his brother are equally responsible for the fights. There is no evidence of substance abuse, domestic violence, or criminal activity by his parents. There is no evidence of any major problems during his childhood (up to the age of 16). However, Mr. Case reports that he often got in trouble for smoking (cigarettes) and staying out late before age 12. Starting at the age of 16, Mr. Case became heavily involved in drug use (see details below).

EDUCATIONAL HISTORY

Mr. Case did not enjoy school and did not do well academically, although before the seventh grade he attended regularly and had a good relationship with peers and teachers. Problems with peers started in the seventh grade. He was slow to mature and was picked on by much bigger students. He was involved in a few minor fights, usually as a result of him being bullied. He never caused a serious injury while in a fight, but was the victim of the occasional black eye. He failed seventh grade, and dropped out in ninth grade "to make money."

EMPLOYMENT HISTORY

After dropping out of school, he worked as a bike messenger for two months before quitting. He quit because he found the work too tiring (biking all day). Mr. Case estimates that he has had between 20 and 25 jobs and that he often became bored and quit after a couple of months. His longest job was for nine months as a dishwasher. He was fired twice for smoking marijuana at work. Mr. Case admits that he was not always a good worker—he would often show up late or miss shifts, and he started one fight with a coworker who he claims insulted him. The coworker received a laceration on his cheek.

Mr. Case had frequent periods of unemployment and typically relied on social assistance. His longest period of unemployment was approximately three years. During this time, Mr. Case hitchhiked his way to Mexico. He did not tell anyone he was leaving, but sent his mother a postcard when he was there. Because of his employment and financial problems, he spent months at a time living on the streets. During his times on the street, Mr. Case relied on his mother for financial support; he would call her and she would send him money. Prior to his admission to SFH, he moved between living with his mother and cheap hotels.

RELATIONSHIP HISTORY

Mr. Case has had five or six short romantic relationships. His longest relationship was for 2 to 3 months, 25 years ago, and he has never lived with a partner. He claims that relationships are not important to him, and he has not expressed any desire to have a relationship in the future.

SUBSTANCE ABUSE HISTORY

Mr. Case started using marijuana, hashish, and LSD when he was 16 years old. Over the next couple of years, Mr. Case used marijuana daily and hash and LSD every weekend and occasionally during the week. At age

19, he stopped using LSD, but started drinking alcohol more often. From the age of 19 to 25, Mr. Case drank alcohol five to six times per week, used marijuana daily, and used hashish three or four times a week. Often, Mr. Case used his social assistance money to pay for drugs. He reports that the drugs were psychologically addictive, and that he liked them because they allowed him "to dream." He claims that that he never did anything dangerous or reckless while using drugs, and there is no evidence to dispute this claim. He has not used LSD for over 25 years, but in the few years prior to his index offense he was drinking five or six days a week and was using marijuana several times per day.

MENTAL HEALTH HISTORY

Mr. Case first had first contact with a mental health professional at age 17 because of drug problems. He only met with his counselor on a few occasions before he stopped attending because he did not think that he had a drug problem.

At the age of 19, he was hospitalized for one month due to a "nervous breakdown." There are few details regarding what occurred to cause the nervous breakdown. He claims that his parents were causing him stress and, combined with his frequent drug and alcohol use, he suffered an emotional breakdown. He was hospitalized again at the age of 25 for two months when his parents were concerned about his frequent yelling at himself. Mr. Case states that he was hearing voices, and he was yelling at these voices to go away. During this hospitalization, staff reported that he was a cooperative patient and easy to manage. Mr. Case's next contact with a mental health professional occurred after his index offense (described below).

Mr. Case has received consistent diagnoses of paranoid schizophrenia and cannabis dependence disorder as a result of several psychodiagnostic assessments conducted over the course of his hospitalization.

CRIMINAL HISTORY

Mr. Case does not have any juvenile convictions or charges. However, he claims that he had some problems with the law before the age of 17, and that he also was charged with assault with a weapon and possession of a restricted weapon prior to his index offense. Mr. Case is unwilling to provide further information, and there are no available records of these alleged offenses because they occurred in Canada.

The index offense occurred May 15, 1982 (at the age of 27), when Mr. Case was living with a man he had met while living on the street. Mr. Case attacked and murdered his roommate while he slept by stabbing him multiple times with a screwdriver and hunting knife. The attack was so severe that the victim was nearly unrecognizable. After the attack, Mr. Case posed the body by laying the victim on his back and crossing his arms on his chest. Following the murder, he stole a number of the victim's possessions, some money, and his car and drove to St. Louis. He was arrested in St. Louis soon after, where the police found him in the victim's car with a bloodstained hunting knife and the victim's possessions (some of which were also bloodstained). When he was arrested, Mr. Case claimed that he did not commit the murder and that the victim gave him his possessions and car and said he would meet him in St. Louis. Mr. Case told the police that the victim was a witch, and that other witches may have done it. When the police confronted him about the bloodstained items, he told them that he did not know how they got there.

Mr. Case was charged with second degree murder, and on June 14, 1982, he was admitted to SFH for a competency evaluation. He was found competent to stand trial and released to court. On October 9, 1982, he was found NGRI for the murder and recommitted to SFH.

ADJUSTMENT AFTER HOSPITALIZATION

In Mr. Case's 26 years at SFH he has been granted conditional discharges to semi-independent living eight times, but has never been able to last more than one year before returning to the hospital. His returns to the hospital were usually a result of his drug use or negative attitudes. He has had many positive drug tests for marijuana that have resulted in his involuntary return to full supervision. On other occasions, he simply did not seem happy while residing at these residences and asked to be returned to the hospital ward. On one occasion, six years ago, he used a knife to severely attack a coresident, whom he accused of being a witch who

was trying to poison him. At the time, he had become noncompliant with medication and was using marijuana heavily. The victim needed 63 stitches about his head, face, and arms, and surgery to stop internal bleeding.

While residing at the hospital, Mr. Case has disobeyed rules on multiple occasions, such as leaving the hospital grounds without permission. On these occasions, he often returned to the downtown area overnight, where he would buy alcohol or marijuana, and voluntarily return to hospital the next day. On other occasions, he would simply return late from his day leaves. Mr. Case often seemed indifferent to having broken rules and reacts defensively when confronted by staff.

Mr. Case's last return to the hospital occurred in April 2005 for breaching his conditions (using alcohol and marijuana). He last went before the Review Board six months ago. The Review Board decided to continue his custody stay because they felt that he had not yet dealt with his substance abuse problems, that he failed to show consistent compliance with treatment and supervision, and that he was at high risk to reoffend as a result of this.

RECENT FUNCTIONING (PAST SIX MONTHS)

Since Mr. Case's last Review Board hearing, he has had a few problems. According to progress reports, he acknowledges that he has a mental illness that he needs treatment for. Mr. Case has stated that using drugs is bad for him, though he denies that his drug use might affect his mental illness, and he continues to use drugs. His privileges were suspended two months ago due to a positive drug screen. He has had four other positive drug screens since his last Review Board hearing.

Mr. Case appears to be compliant with his medications, and reports indicate that he is responding well. Recent staff reports have indicated that there have been occasional instances where he has delusional/disorganized thinking, odd movements, and has been smiling and laughing to himself, but that these behaviors have become less frequent in the past four months. In an interview two months ago, Mr. Case reported that when he takes drugs he "sees the wonders of the universe and can understand the mysteries of the world." He also reported that he "has met God and can enter the spirit world." His treatment team feels that his mental functioning is significantly affected by his substance use, and his continued use of marijuana has been his biggest problem.

Mr. Case claims that his treatment has been helpful, although he has not participated in the drug and alcohol program as recommended. Instead, he relies on a weekly meeting with his pastor for assistance. He states that this is the only person he trusts at the hospital. He often complains about the staff, states that he does not trust them, and feels that they are not "too together."

There has been the occasional minor conflict with other residents at the hospital, usually involving a verbal disagreement. The situation has not escalated, and according to staff reports, Mr. Case has handled these situations well. If he does feel angry, he often will take a short walk around the hospital grounds to calm himself down and/or speak with his pastor for advice. These two strategies have been effective for him according to treatment staff and his self-report.

For the past six months, Mr. Case has been working in the hospital recycling plant three days a week. His supervisor reports that he is a good worker: he is always on time, completes the tasks assigned to him, has only missed work due to legitimate reasons, and gets along well with his coworkers. On the days he does not work he often attends a drop-in center near the downtown area. At the moment, he is not permitted to attend this center because the hospital staff believe it is after these trips to the drop-in center, while he is still downtown, that he buys and uses drugs. His remaining leisure time is spent practicing yoga and reading books, mainly on the topic of Christianity.

RELEASE PLANS

Mr. Case currently has no plans as to where he would live if released. He also does not know where, or if, he will work. If he is required to find a new job, he is not sure of what type he would like and plans to rely solely on his disability benefits if he cannot find any employment that he wants. When asked about his treatment plan after release, he replies that his treatment will be up to his doctors. If it were up to him, he would not continue with treatment. When asked how he will keep himself safe, he states that he will "keep

his house clean and mind his own business." If his symptoms start to worsen, he reports that he will seek help, but he did not explain from where he will get this help. He stated that, if the doctors recommend that he continue with his medications, he will, and he doesn't think he will have any problems with compliance. Mr. Case's treatment team has encouraged him to make more substantial plans for his release, but he has not yet complied.

Mr. Case is not concerned about facing any destabilizers. When asked specifically about the risk of using drugs and alcohol, he reports that he will quit drugs on his own, but cannot elaborate on any plan for this and says he will "play it by ear." His treatment team is concerned about his attitude toward his drug and alcohol use and has strongly suggested a community treatment program. Mr. Case has replied that, if the program is a required condition of his release, he will attend a program, but if it is only a "suggestion" he does not know if he will attend.

In terms of possible sources of personal support, both of his parents have passed away, and he has not been in touch with his brother since his mother's death ten years ago. He has made some friends while in the hospital, but he does not expect to keep in touch with them when released and no longer has friends in the community due to the length of time he has been hospitalized. His one source of support is his pastor. The treatment team is familiar with Mr. Case's pastor and believes that he will be a good source of support for Mr. Case.

Mr. Case does not expect to face any stressors when released. Although he does not have any plans for where he will live or how he will make money, he says that things will work themselves out. He feels that he handles stress well and that he will be able to handle any situation that he is faced with. Mr. Case does not feel that he is at risk to commit a new crime when he is released.

RISK ASSESSMENT

In order to inform my judgment of risk, I used the HCR-20 (Historical-Clinical-Risk Management 20) violence risk assessment instrument. The HCR-20 has been subjected to more than 50 empirical studies that, on average, indicate that it facilitates consistency in clinical evaluations of risk for violence, and that its risk factors and judgments of risk based upon them help evaluators to determine whether an individual poses a low, moderate, or high level of future risk for violence. This HCR-20 includes 20 risk factors to consider, along with standard instructions for how to rate them; I determined whether each of these risk factors was present for Mr. Case, and if so, how relevant it appeared to be for Mr. Case's risk of violence.

The HCR-20 includes 10 Historical risk factors—past events, experiences, or conditions that are known to increase the risk for violence. I determined that eight of these risk factors are definitely present for Mr. Case: previous violence (H1), young age at first violent incident (H2), relationship instability (H3), employment problems (H4), substance use problems (H5), major mental illness (H6), early maladjustment (H8), and prior supervision failure (H10). Two other risk factors from this scale were possibly or partially present: psychopathy (H7) and personality disorder (H9).

In terms of recent functioning, the HCR-20 includes five risk factors on its Clinical Scale that are intended to summarize a person's recent emotional, behavioral, and cognitive functioning. I determined whether these risk factors were present within the past six months—since Mr. Case's last Review Board hearing. I determined that three of the five risk factors definitely were present within the past six months: lack of insight (C1), active symptoms of major mental illness (C3), and unresponsive to treatment (C5). In addition, the other two risk factors were possibly or partially present: negative attitudes (C2) and impulsivity (C4).

The HCR-20 Risk Management Scale includes five risk factors that pertain to an individual's possible future living circumstances. I rated these risk factors as if Mr. Case were to be released into the community, rather than for continued hospitalization. That is, I determined, given Mr. Case's current release planning, whether the risk factors would be present if Mr. Case were released into the community by the Review Board next month. I used a six-month time frame for these ratings.

I determined that four of the five Risk Management risk factors were present for Mr. Case: plans lack feasibility (R1), exposure to destabilizers (R2), lack of personal support (R3), and noncompliance with remediation attempts (R4). One of the risk factors was possibly present—stress (R5).

Therefore, there are 15 of the HCR-20's 20 risk factors that definitely present for Mr. Case, and another 5 that are possibly or partially present. Some of these 15 present risk factors are especially relevant to understanding Mr. Case's risk for future violence, as follows.

There are several clusters of risk factors are of particular relevance and concern and also seem likely to influence and worsen one another. First, Mr. Case has paranoid schizophrenia (H6), and there is evidence of active psychotic symptoms within the past six months (C3). The murder that Mr. Case committed (H1) occurred when Mr. Case was experiencing psychosis, as did the more recent attack on a coresident; both appear to have been motivated by his psychotic symptoms. Second, Mr. Case has frequently been noncompliant with past supervision and risk management efforts (H10), has recently been noncompliant with supervision orders pertaining to drug use (C5), and has expressed considerable ambivalence about whether he will comply with treatment in the future (R4). Third, Mr. Case has a serious substance dependence disorder (H5) that he fails to acknowledge (C1) or seek treatment for (C5) and which appears to exacerbate his psychotic symptoms. Fourth, if Mr. Case were released from the hospital, especially without conditions, he likely would be exposed to situations and contexts (such as drug availability), which would destabilize his mental health (R2). In addition, he has not developed any coherent discharge plans (R1) that would provide him with stable housing, effective risk management and treatment, or reliable income.

There are two further areas of concern. First, Mr. Case has very little social support from friends, family, or a romantic partner (R3, H3). Such support can mitigate risk, but is limited for Mr. Case, other than his relationship with his pastor. Another potential risk-reducing opportunity—stable employment—also is missing (H4). Although he has worked effectively of late while in the institution, he has a clear record of employment problems when he has had the opportunity to work in the community, and he has no plans to work after release. The second main area of concern is Mr. Case's antisocial or psychopathic traits (H7, H9) and distrust of most professionals who work with him. This may make securing his agreement to start, comply with, and complete programming more challenging.

Given both the presence of a large number of risk factors and their relevance to Mr. Case's risk for violence, in my judgment Mr. Case is a high risk for violence. Furthermore, although history is not destined to repeat itself, consideration of H1 (previous violence) indicates the type of violence that might be of concern if Mr. Case were to act violently. That is, in my judgment Mr. Case is at high risk for violence that could cause serious injuries to others.

A judgment of high risk means that (1) there is a high likelihood that, if released soon with the current discharge plans, Mr. Case will act violently within the next six months; (2) Mr. Case should be considered high priority for the delivery of supervision and management resources; and (3) Mr. Case requires a high level or intensity of supervision and management in order to mitigate risk. While it is not possible to produce a meaningful numeric estimate of risk, there are numerous risk factors that have not been controlled, and for which there are no future risk management plans in place. I would recommend further efforts to engage Mr. Case in drug and alcohol treatment, and to increase his compliance with supervision and treatment efforts, perhaps through establishing a set of contingencies for compliance and noncompliance. It is also important to monitor his psychotic symptoms closely. It may also be worthwhile to foster Mr. Case's recent employment success and to assist him in finding employment and stable housing. If Mr. Case were to be released, I would recommend frequent (biweekly) monitoring and reevaluation of his risk factors.

CASE COMMENT

This case example provides an illustration of one potential application of the HCR-20, and one example of how to format an HCR-20 risk assessment report. Further, if this were an actual case report, evaluators might want to spend more time describing the details of risk management plans, which space limitations precluded us from doing. The important point that we were illustrating was that judgments of risk should be grounded in the case-specific evidence (e.g., the individual manifestation and relevance of HCR-20 risk factors, which themselves have nomothetic support), and that recommendations for risk management should be logically linked to the risk factors of most concern.

Historical-Clinical-Risk Management-20 (HCR-20) Violence Risk Assessment Scheme • **181**

References

Allen, C., & Howells, K. (2008). *The implementation and evaluation of a structured professional judgment risk assessment tool within a high secure forensic hospital.* Paper presented at the annual conference of the International Association of Forensic Mental Health Services, Vienna, Austria.

Andrews, D. A., Bonta, J., & Wormith, J. (2004). *The Level of Service/Case Management Inventory user's manual.* North Tonawanda, NY: Multi-Health Systems.

Belfrage, H. (1998). Implementing the HCR-20 scheme for risk assessment in a forensic psychiatric hospital: Integrating research and clinical practice. *Journal of Forensic Psychiatry, 9,* 328–338.

Belfrage, H., & Douglas, K. S. (2002). Treatment effects on forensic psychiatric patients measured with the HCR-20 violence risk assessment scheme. *International Journal of Forensic Mental Health, 1,* 25–36.

Blair, P. R., Marcus, D. K., & Boccaccini, M. T. (2008). Is there an allegiance effect for assessment instruments? Actuarial risk assessment as an exemplar. *Clinical Psychology: Research and Practice, 15,* 346–360.

Blum, F. M. (2004). *Psychopathy, psychosis, drug abuse, and reoffense among conditionally released offenders.* Unpublished doctoral dissertation, University of Southern California, Los Angeles, CA.

Brown, L. K. (2004). *Assessing risk for elopement and breaches of conditional release in insanity acquittees.* Unpublished doctoral dissertation, Simon Fraser University, Burnaby, BC, Canada.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86,* 127–137.

Claix, A., Pham, T., & Willocq, L. (2002, March). *Evaluation of the HCR-20 (historical-clinical-risk management) in a Belgian forensic population.* Poster presented at the annual conference of the international association of forensic mental health services, Munich, Germany.

Cooke, D. J., Kosson, D. S., & Michie, C. (2001). Psychopathy and ethnicity: Structural, item, and test generalizability of the Psychopathy Checklist—Revised (PCL-R) in Caucasian and African American participants. *Psychological Assessment, 13,* 531–542.

Cooke, D. J., & Michie, C., & Ryan, J. (2001). Evaluating Risk for Violence: A Preliminary Study of the HCR-20, PCL-R and VRAG in a Scottish Prison Sample. Scottish Prison Service occasional papers; no 5/200.

Côté, G., Hodgins, S., & Daigle, M. (2001). *Violent behaviour, PCL-R and HCR-20 among involuntary inpatients, forensic patients and severely mentally disordered inmates.* Paper presented at the First Annual Meeting of the International Association of Forensic Mental Health Services, Vancouver, British Columbia.

Dahle, K. (2006). Strengths and limitations of actuarial prediction of criminal reoffence in a German prison sample: A comparative study of LSI-R, HCR-20 and PCL-R. *International Journal of Law and Psychiatry, 29,* 431–442.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34,* 571–582.

de Vogel, V., & de Ruiter, C. (2005). The HCR-20 in personality disordered female offenders: A comparison with a matched sample of males. *Clinical Psychology & Psychotherapy, 12,* 226–240.

de Vogel, V., & de Ruiter, C. (2006). Structured professional judgment of violence risk in forensic clinical practice: A prospective study into the predictive validity of the Dutch HCR-20. *Psychology, Crime & Law, 12,* 321–336.

de Vogel, V., de Ruiter, C., Hildebrand, M., Bos, B., & van de Ven, P. (2004). Type of discharge and risk of recidivism measured by the HCR-20: A retrospective study in a Dutch sample of treated forensic psychiatric patients. *International Journal of Forensic Mental Health, 3,* 149–165.

Dernevik, M. (1998). Preliminary findings on reliability and validity of the Historical-Clinical-Risk Assessment in a forensic psychiatric setting. *Psychology, Crime, and Law, 4,* 127–137.

Dernevik, M., Grann, M., & Johansson, S. (2002). Violent behaviour in forensic psychiatric patients: Risk assessment and different risk-management levels using the HCR-20. *Psychology, Crime & Law, 8,* 93–111.

Dolan, M., & Fullam, R. (2007). The validity of the Violence Risk Scale 2nd edition (VRS-2) in a British Forensic Inpatient Sample. *The Journal of Forensic Psychology and Psychiatry, 18,* 381–193.

Dolan, M., & Khawaja, A. (2004). The HCR-20 and post-discharge outcome in male patients discharged from medium security in the UK. *Aggressive Behavior, 30,* 469–483.

Doren, D. M. (2002). *Evaluating sex offenders: A manual for civil commitments and beyond.* Thousand Oaks, CA: Sage.

Douglas, K. S. (2008). The HCR-20 violence risk assessment scheme. In B. Cutler (Ed.), *Encyclopedia of psychology and law.* Thousand Oaks, CA: Sage.

Douglas, K. S., & Belfrage, H. (2001). Use of the HCR-20 in violence risk management: Implementation and clinical practice. In K. S. Douglas, C. D. Webster, S. D. Hart, D. Eaves, & J. R. P. Ogloff, (Eds.), *HCR-20: Violence risk management companion guide* (pp. 41–58). Burnaby, BC, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

Douglas, K. S., Guy, L. S., Reeves, K. A., & Weir, J. (2008). *HCR-20 violence risk assessment scheme: Overview and annotated bibliography.* Retrieved April 5, 2009, from http://kdouglas.files.wordpress.com/2006/04/annotate10-24nov2008.pdf

Douglas, K. S., Hart, S. D., Webster, C. D., Belfrage, H., & Eaves, D. (2008). *HCR-20: Assessing Risk for Violence* (Version 3, Draft 1.0 Nov. 14, 2008). Burnaby, BC, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

Douglas, K. S., & Kropp, P. R. (2002). A prevention-based paradigm for violence risk assessment: Clinical and research applications. *Criminal Justice and Behavior, 29,* 617–658.

Douglas, K. S., Nicholls, T. L., & Brink, J. (2009). Reducing the risk of violence among persons with mental illness: A critical analysis of treatment approaches. In P. M. Kleespies (Ed.), *Evaluating and managing behavioral emergencies: An evidence-based resource for the mental health practitioner* (pp. 351–376). Washington, DC: American Psychological Association.

Douglas, K. S., Ogloff, J. R. P., & Hart, S. D. (2003). Evaluation of a model of violence risk assessment among forensic psychiatric patients. *Psychiatric Services, 54,* 1372–1379.

Douglas, K. S., Ogloff, J. R. P., Nicholls, T. L., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 violence risk assessment scheme and the Psychopathy Checklist: Screening Version. *Journal of Consulting and Clinical Psychology, 67,* 917–930.

Douglas, K., S., & Skeem, J. L. (2005). Violence risk assessment: Getting specific about being dynamic. *Psychology, Public Policy, and Law, 11,* 347–383.

Douglas, K. S., Webster, C. D., Hart, S. D., Eaves, D., & Ogloff, J. R. P. (Eds.) (2001). HCR-20: *Violence risk management companion guide.* Burnaby, BC, Canada: Mental Health, Law, and Policy Institute, Simon Fraser University.

Douglas, K. S., Yeomans, M., & Boer, D. P. (2005). Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior, 32,* 479–510.

Dowden, C., & Andrews, D. A. (2000). Effective correctional treatment and violent reoffending: A meta-analysis. *Canadian Journal of Criminology, 42,* 449–467.

Doyle, M., & Dolan, M. (2006). Predicting community violence from patients discharged from community mental health services. *British Journal of Psychiatry, 189,* 520–526.

Doyle, M., Dolan, M., & McGovern, J. (2002). The validity of North American risk assessment tools in predicting in-patient violent behaviour in England. *Legal and Criminological Psychology, 7,* 141–152.

Dunbar, E. (2003). Symbolic, relational, and ideological signifiers of bias-motivated offenders: Toward a strategy of assessment. *American Journal of Orthopsychiatry, 73,* 203.

Dunbar, E., Quinones, J., & Crevecoeur, D. A. (2005). Assessment of hate crime offenders: The role of bias intent in examining violence risk. *Journal of Forensic Psychology Practice, 5,* 1–19.

Dvoskin, J. A., & Heilbrun, K. (2001). Risk assessment and release decision-making: Toward resolving the great debate. *Journal of the American Academy of Psychiatry and the Law, 29,* 6–10.

Fujii, D., Lichton, A., & Tokioka, A. (2004, August). *Structured professional judgment versus actuarial data in violence risk prediction using the Historical Clinical Risk Management-20.* Paper presented at the annual convention of the American Psychological Association, Honolulu, HI.

Fujii, D. E. M., Tokioka, A. B., Lichton, A. I., & Hishinuma, E. (2005). Ethnic differences in prediction of violence risk with the HCR-20 among psychiatric inpatients. *Psychiatric Services, 56,* 711–716.

Grann, M., Belfrage, H., & Tengstrom, A. (2000). Actuarial assessment of risk for violence: Predictive validity of the VRAG and the historical part of the HCR-20. *Criminal Justice and Behavior, 27,* 97–114.

Grann, M., & Långström, N. (2007). Actuarial assessment of violence risk: To weigh or not to weigh? *Criminal Justice and Behavior, 34,* 22–36.

Grann, M., Sturidsson, K., Haggard-Grann, U., Hiscoke, U. L., Alm, P., & Dernevik, M., et al. (2005). Methodological development: Structured outcome assessment and community risk monitoring (SORM). *International Journal of Law and Psychiatry, 28,* 442–456.

Grann, M., & Wedin, I. (2002). Risk factors for recidivism among spousal assault and spousal homicide offenders. *Psychology, Crime and Law, 8,* 5–23.

Gray, N. S., Fitzgerald, S., Taylor, J., MacCulloch, M. J., & Snowden, R. J. (2007). Predicting future reconviction in offenders with intellectual disabilities: The predictive efficacy of VRAG, PCL-SV and the HCR-20. *Psychological Assessment, 19,* 474–479.

Gray, N. S., Hill, C., McGleish, A., Timmons, D., MacCulloch, M. J., & Snowden, R. J. (2003). Prediction of violence and self-harm in mentally disordered offenders: A prospective study of the efficacy of HCR-20, PCL-R, and psychiatric symptomatology. *Journal of Consulting and Clinical Psychology, 71*, 443–451.

Gray, N. S., Snowden, R. J., MacCulloch, S., Phillips, H., Taylor, J., & MacCulloch, M. J. (2004). Relative efficacy of criminological, clinical, and personality measures of future risk of offending in mentally disordered offenders: A comparative study of HCR-20, PCL:SV, and OGRS. *Journal of Consulting and Clinical Psychology, 72*, 523–530.

Gray, N. S., Taylor, J., & Snowden, R. J. (2008) Predicting violent reconvictions using the HCR-20. *British Journal of Psychiatry: The Journal of Mental Science, 192*, 384–387.

Grevatt, M., Thomas-Peter, B., & Hughes, G. (2004). Violence, mental disorder and risk assessment: Can structured clinical assessments predict the short-term risk of inpatient violence? *Journal of Forensic Psychiatry & Psychology, 15*, 278–292.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law, 2*, 293–323.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19–30.

Guy, L. S. (2008). *Performance indicators of the structured professional judgment approach for assessing risk for violence to others: A meta-analytic survey.* Unpublished dissertation, Simon Fraser University, Burnaby, BC, Canada.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior, 20*, 315–335.

Hart, S. D. (1998). The role of psychopathy in assessing risk for violence: Conceptual and methodological issues. *Legal and Criminological Psychology, 3*, 121–137.

Hartvig, P., Alfarnes, S. A., Skjonberg, M., Moger, T. A., & Ostberg, B. (2006). Brief checklists for assessing violence risk among patients discharged from acute psychiatric facilities: A preliminary study. *Nordic Journal of Psychiatry, 60*, 243–248.

Heilbrun, K., Douglas, K. S, & Yasuhara, K. (2009). Controversies in violence risk assessment. In J. L. Skeem, K. S. Douglas, & S. O. Lilienfeld (Eds), *Psychological science in the courtroom: Controversies and consensus* (pp. 333–357). New York: Guilford Press.

Hildebrand, M., Hesper, B. L., Spreen, M. & Nijman, H. L.I. (2005). *De waarde van gestructureerde risicotaxatie en van de diagnose psychopathie: een onderzoek naar de betrouwbaarheid en predictieve validiteit van de HCR-20, HKT-30 en 195 PCL-R.* [The importance of structured risk assessment and of the diagnosis of psychopathy: A study into the reliability and predictive validity of the HCR-20, HKT-30 and PCL-R]. Utrecht, the Netherlands: Expertisecentrum Forensische Psychiatrie.

Hill, A., Habermann, N., Klusmann, D., Berner, W., & Briken, P. (2008). Criminal recidivism in sexual homicide perpetrators. *International Journal of Offender Therapy and Comparative Criminology, 52*, 5–20.

Holtfreter, K., & Cupp, R. (2007). Gender and risk assessment: The empirical status of the LSI-R for women. *Journal of Contemporary Criminal Justice, 23*, 363–382.

Howard, B. (2007). *Examining predictive validity of the Salient Factor Score and HCR-20 among behavior health court clientele: Comparing static and dynamic variables.* Unpublished doctoral dissertation, Pacific Graduate School of Psychology, Palo Alto, CA.

Howells, K., Watt, B., Hall, G., & Baldwin, S. (1997). Developing programmes for violent offenders. *Legal and Criminological Psychology, 2*, 117–128.

Kroner, D. G., & Mills, J. F. (2001). The accuracy of five risk appraisal instruments in predicting institutional misconduct and new convictions. *Criminal Justice and Behavior, 28*, 471–489.

Landis, J., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Laws, D. R., & O'Donohue, W. T (Eds.). (2008). *Sexual deviance: Theory, assessment, and treatment* (2nd ed.). New York: Guilford Press.

Lindsay, W., Hogue, T., Taylor, J., Steptoe, L., Mooney, P., O'Brien, G., et al. (2008). Risk assessment in offenders with intellectual disability: A comparison across 201 three levels of security. *International Journal of Offender Therapy and Comparative Criminology, 52*, 90–111.

MacPherson, G. J. D., & Kevan, I. (2004). Predictive validity of the HCR-20 violence risk assessment scheme within a maximum security special hospital. *Issues in Forensic Psychology, 5*, 62–80.

McDermott, B. E., Edens, J. F., Quanbeck, C. D., Busse, D., & Scott, C. L. (2008). Examining the role of static and dynamic risk factors in the prediction of inpatient violence: Variable- and person-focused analyses. *Law and Human Behavior, 34*, 325–338.

McEachran, A. K. (2001). *The predictive validity of the PCL:YV and the SAVRY in a population of adolescent offenders.* Unpublished master's thesis, Simon Fraser University, Burnaby, BC, Canada.

McKenzie, B., & Curr, H. (2005). Predicting violence in a medium secure setting: A study using the historical and clinical scales of the HCR-20. *British Journal of Forensic Practice, 7,* 22–28.

McNiel, D. E., Gregory, A. L., Lam, J. N., Binder, R. L., & Sullivan, G. R. (2003). Utility of decision support tools for assessing acute risk of violence. *Journal of Consulting & Clinical Psychology, 71,* 945–953.

Meehl, P. E. (1954). *Clinical versus statistical prediction.* Minneapolis, MN: University of Minnesota Press.

Mills, J. F., Jones, M. N., & Kroner, D. G. (2005). An examination of the generalizability of the LSI-R and VRAG probability bins. *Criminal Justice and Behavior, 32,* 565–585.

Mills, J., Kroner, D., & Hemmati, T. (2007). The validity of violence risk estimates: An issue of item performance. *Psychological Services, 4,* 1–12.

Monahan, J., Steadman, H. J., Appelbaum, P. S., Grisso, T., Mulvey, E. P., Roth, L. H., Robbins, P. C., Banks, S., & Silver, S. (2005). *Classification of Violence Risk (COVR).* Lutz, FL: Psychological Assessment Resources.

Morrissey, C., Hogue, T., Mooney, P., Allen, C., Johnston, S., Hollin, C., Lindsay, W. R,. & Taylor, J. L. (2007). Predictive validity of the PCL-R in offenders with intellectual disability in a high secure hospital setting: Institutional aggression. *Journal of Forensic Psychiatry & Psychology, 18,* 1–15.

Mossman, D., & Somoza, E. (1991). ROC curves, test accuracy, and the description of diagnostic tests. *Journal of Neuropsychiatry and Clinical Neurosciences, 3,* 330–333.

Müller-Isberner, J. R., Sommer, J., Özokyay, K., & Freese, R. (1999, November). *Clinical use of the HCR-20 for predicting violence in a German forensic psychiatric hospital.* Paper presented at the International Conference on Risk Assessment and Management: Implications for Prevention of Violence, Vancouver, BC, Canada.

Mulvey, E. P., & Lidz, C. W. (1995). Conditional prediction: A model for research on dangerousness to others in a new era. *International Journal of Law and Psychiatry, 18,* 129–143.

Neves, A. C., & Gonçalves, R. A. (July, 2008). *Criminal recidivism and violation of conditional release: A comparative study of the LSI-R and the HCR-20 with Portuguese Probationers/parolees.* Poster presented at the annual conference of the International Association of Forensic Mental Health Services, Vienna, Austria.

Nicholls, T. L. (2001). *Violence risk assessments with female NCRMD acquittees: Validity of the HCR-20 and PCL-SV.* Unpublished doctoral dissertation, Simon Fraser University, Burnaby, BC, Canada.

Nicholls, T. L., Ogloff, J. R. P., & Douglas, K. S. (2004). Assessing risk for violence among female and male civil psychiatric patients: The HCR-20, PCL: SV, and McNiel & Binder's VSC. *Behavioral Sciences and the Law, 22,* 127–158.

Pham, T. H., Claix, A., & Remy, S. (2000, June). *Assessment of the HCR-20 in a Belgian prison sample.* Paper presented at the 4th European Congress on Personality Disorders. Paris, France.

Pham, T. H., Ducro, C., Marghem, B., & Réveillère, C. (2005). Évaluation du risque de récidive au sein d'une population de délinquants incarcérés ou internés en Belgique francophone [Prediction of recidivism among prison inmates and forensic patients in Belgium]. *Annales Médico-Psychologiques, 163,* 842–845.

Philipse, M. (2002, March). *Post-dictive validity of the HCR-20 in a Dutch forensic psychiatric sample.* Paper presented at the Annual Conference of the International Association of Forensic Mental Health Services, Munich, Germany.

Polvi, N. H. (1999). *The prediction of violence in pre-trial forensic patients: The relative efficacy of statistical versus clinical predictions of dangerousness.* Unpublished doctoral dissertation, Simon Fraser University, Burnaby, BC, Canada.

Quinsey, V. L., Harris, G. T., Rice, G. T., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk.* Washington, DC: American Psychological Association.

Quinsey, V. L., Harris, G. T., Rice, G. T., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). Washington, DC: American Psychological Association.

R. v. McCraw, (1991). 3 S.C.R. 72.

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC, Cohen's d and r. *Law and Human Behavior, 29,* 615–620.

Ross, D. J., Hart, S. D., & Webster, C. D. (1998). *Facts and fates: Prediction and management of hospital and community aggression using the HCR-20.* Port Coquitlam, British Columbia: Riverview Hospital.

Skeem, J., & Mulvey, E. (2002). Monitoring the violence potential of mentally disordered offenders being treated in the community. In A. Buchanan (Ed.), *Care of the mentally disordered offender in the community* (pp. 111–142). New York: Oxford Press.

Stadtland, C., Hollweg, M., Kleindienst, N., Dietl, J., Reich, U., & Nedopil, N. (2005). Risk assessment and prediction of violent and sexual recidivism in sex offenders: Long-term predictive validity of four risk assessment instruments. *Journal of Forensic Psychiatry & Psychology, 16,* 92–108.

Stadtland, C., & Nedopil, N. (2005). [Psychiatric disorders and the prognosis for criminal recidivism]. *Der Nervenarzt, 76,* 1402–1411.

Strand, S., & Belfrage, H. (2001). Comparison of HCR-20 scores in violent mentally disordered men and women: Gender differences and similarities. *Psychology, Crime and Law, 7,* 71–79.

Strand, S., Belfrage, H., Fransson, G., & Levander, S. (1999). Clinical and risk management factors in risk prediction of mentally disordered offenders—MORE important than historical data? A retrospective study of 40 mentally disordered offenders assessed with the HCR-20 violence risk assessment scheme. *Legal and Criminological Psychology, 4,* 67–76.

Tengström, A. (2001). Long-term predictive validity of historical factors in two risk assessment instruments in a group of violent offenders with schizophrenia. *Nordic Journal of Psychiatry, 55,* 243–249.

Tengström, A., Hodgins, S., Müller-Isberner, R., Jockel, D., Freese, R., & Ozokyay, K., et al. (2006). Predicting violent and antisocial behavior in hospital using the HCR-20: The effect of diagnoses on predictive accuracy. *International Journal of Forensic Mental Health, 5,* 39–53.

Urheim, R., Jakobsen, D., & Rasmussen, K. (2003, August). *Dimensions of inpatient aggressive behavior in a security ward: What is being "predicted"?* Paper presented at the 5th Nordic Symposium on Forensic Psychiatry, Ystad, Sweden.

Warren, J. I., South, S. C., Burnette, M. L., Rogers, A., Friend, R., Bale, R., Van Patten, I. (2005). Understanding the risk factors for violence and criminality in women: The concurrent validity of the PCL-R and HCR-20. *International Journal of Law and Psychiatry, 28,* 269–289.

Webster, C. D., Douglas, K., Eaves, D. & Hart, S. (1997). *HCR-20: Assessing risk for violence,* Version 2. Burnaby, BC, Canada: Simon Fraser University.

Webster, C. D., Eaves, D., Douglas, K. S., & Wintrup, A. (1995). *The HCR-20 scheme: The assessment of dangerousness and risk.* Burnaby, BC, Canada: Simon Fraser University and British Columbia Forensic Psychiatric Services Commission.

Wintrup, A. (1996). *Assessing risk of violence in mentally disordered offenders with the HCR-20. Unpublished master's thesis,* Simon Fraser University, Burnaby, BC, Canada.